

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 6, 2001, 20:59:36 : Search time 51.66 Seconds  
(without alignments)  
726.850 Million cell updates/sec

Title: US-09-292-862-2

Perfect score: 553  
Sequence: 1 MQARYSVSPNSLGVVPLIG.....PSSQSLYRTSGAFVYDCSKF 553

Scoring table: OLIGO  
Gapop 60.0, Gapext 60.0

Searched: 195891 seqs, 67900655 residues

Word size : 0

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: PIR66:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	109	19.7	117 2	149674
2	106	19.2	106 2	FS1626
3	61	11.0	461 2	FS4472
4	59	10.7	101 2	160918
5	39	7.1	108 2	B48924
6	35	6.3	111 2	B56556
7	25	4.5	508 2	S59870
8	23	4.2	111 2	C47746
9	22	4.0	111 2	F47746
10	21	3.8	270 2	T16880
11	19	3.4	445 2	S23055
12	19	3.4	445 2	S23056
13	16	2.9	106 2	SS1630
14	16	2.9	109 2	B54743
15	16	2.9	134 2	B46178
16	16	2.9	387 2	A47446
17	16	2.9	451 2	A55909
18	16	2.9	469 2	A57451
19	16	2.9	476 2	A54743
20	16	2.9	480 2	JH0672
21	15	2.7	101 2	160922
22	15	2.7	101 2	160917
23	15	2.7	106 2	S51627
24	15	2.7	128 2	C46178
25	15	2.7	202 2	T30169
26	15	2.7	202 2	T15311
27	15	2.7	440 2	S71795
28	15	2.7	465 2	G02738
29	14	2.5	101 2	160920

30	14	2.5	111 2	B47746
31	13	2.4	76 2	S51629
32	13	2.4	101 2	A47450
33	13	2.4	101 2	160919
34	13	2.4	111 2	G56556
35	13	2.4	322 2	S23053
36	13	2.4	367 2	S49008
37	13	2.4	370 2	S49008
38	13	2.4	443 1	S29354
39	13	2.4	445 1	S31224
40	13	2.4	517 2	T43358
41	13	2.4	543 2	F70726
42	13	2.4	543 2	A41285
43	13	2.4	586 2	JG6501
44	13	2.4	617 2	A56051
45	13	2.4	642 2	T39607

#### ALIGNMENTS

##### RESULT 1

149674 forkhead transcription factor fkh-1 - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 02-Jul-1996 #sequence-revision 02-Jul-1996 #text-change 24-Sep-1999

C:Accession: 149674; A47746; S36070

R:Saakl, H.; Hogan, B.L., 1993

Development 118, 47-59, 1993

A:Title: Differential expression of multiple fork head related genes during gastrulation

A:Reference number: 149674; MUID:93387221

A:Accession: 149674

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Accession: 149674

A:Reference number: 149674; MUID:93387221

A:Reference number: A47746; MUID:93387221

A:Accession: A47746

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 109, 1-109 <KAE>

A:Cross-references: EMBL:X71939, NID:g311736; PIDN:CAA50741.1; PID:g311737

A:Note: submitted to the EMBL data library, May 1993

C:Note: sequence extracted from NCBI backbone (NCBI:P136529)

C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology

C:Keywords: DNA binding; transcription factor

F:8-99/Domain: fork head DNA-binding domain homology <FHD>

Query Match

Best Local Similarity 100.0%; Score 109; DB 2; Length 117;

Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 POPKDMVPPSYALITMAIONAPDKITLNGIYOFIMDRPEPTRDNKGOMNIRNL 130

DB 1 POPKDMVPPSYALITMAIONAPDKITLNGIYOFIMDRPEPTRDNKGOMNIRNL 60

QY 131 SLNCEVAVPRDDKPKGSGYWTLDPSYNNFENGSLRRRRFKKDA 179

DB 61 SLNCEVAVPRDDKPKGSGYWTLDPSYNNFENGSLRRRRFKKDA 109

RESULT 2

FS1626

FS1626

FS1626

FS1626

FS1626

FS1626

FS1626

FS1626

FS1626

FS1626

A:Title: Cloning and characterization of seven human forkhead proteins: binding site spd  
A:Reference number: S51624; MUID:95045392

A:Accession: S51626

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-106 <PIE>

A:Cross-references: EMBL:U13221; NID:9563161; PIDN:AAA2038.1; PID:9563162

C:Genetics:

A:Gene: GDB:FKHL7; FREAC3

A:Cross-references: GDB:450220; OMIM:601090

A:Map position: 6p25-6p25

C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology

F:6-97/Domain: fork head DNA-binding domain homology <FHD>

Query Match

Best Local Similarity 19.2%; Score 106; DB 2; Length 106;

Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 PKDMKPPSYTALITMAIQNAPDKKITLNGIYQFIMDRPEFYDNKGQNSIRHNSL 132

Db 1 PKDMKPPSYTALITMAIQNAPDKKITLNGIYQFIMDRPEFYDNKGQNSIRHNSL 60

QY 133 NECFVKVPRDDKPKGKSGYWTLPDPSYMFENGSLRRRRRKKD 178

Db 61 NECFVKVPRDDKPKGKSGYWTLPDPSYMFENGSLRRRRRKKD 106

RESULT 3

S34472

MFH-1 protein - mouse

C:Species: Mus musculus (house mouse)

C>Date: 02-Dec-1993 #sequence\_revision 01-Sep-1995 #text\_change 24-Sep-1999

C:Accession: S34472

R:Miura, N.; Wanka, A.; Tohyama, M.; Tanaka, K.

FEBS Lett. 326, 171-176, 1993

A:Title: MFH-1, a new member of the fork head domain family, is expressed in developing

A:Reference number: S34472; MUID:93314779

A:Accession: S34472

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-461 <MID>

A:Cross-references: GB:S63607; NID:9386637; PIDN:AAB27463.1; PID:9386638

C:Genetics:

A:Gene: MFH-1

C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology

F:38-129/Domain: fork head DNA-binding domain homology <FHD>

Query Match

Best Local Similarity 11.0%; Score 61; DB 2; Length 461;

Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 NKQGMQNSIRHNSLNECFVKVPRDDKPKGKSGYWTLPDPSYMFENGSLRRRRRKK 177

Db 78 NKQGMQNSIRHNSLNECFVKVPRDDKPKGKSGYWTLPDPSYMFENGSLRRRRRKK 137

QY 178 D 178

Db 138 D 138

RESULT 4

160918

Brain factor-3 - rat (fragment)

C:Species: Rattus norvegicus (Norway rat)

C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 24-Sep-1999

C:Accession: 160918

R:Cleveland, D.E.; Overdier, D.G.; Tao, W.; Qian, X.; Paul, L.; Lai, E.; Costa, R.H.

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-101 <RES>

A:Cross-references: GB:LI3193; NID:9310156; PIDN:AAA41320.1; PID:9310157

C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology

F:3-94/Domain: fork head DNA-binding domain homology <FHD>

Query Match

Best Local Similarity 10.7%; Score 59; DB 2; Length 101;

Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 NKQGMQNSIRHNSLNECFVKVPRDDKPKGKSGYWTLPDPSYMFENGSLRRRRRKK 176

Db 43 NKQGMQNSIRHNSLNECFVKVPRDDKPKGKSGYWTLPDPSYMFENGSLRRRRRKK 101

RESULT 5

B48924

forkhead transcription activator homolog (clone FKH H8) - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 21-Jan-1994 #sequence\_revision 18-Nov-1994 #text\_change 12-Sep-1997

C:Accession: B48924

R:Thomas, R.; Moore, J.; Johnston, T.; Socha, C.; Klemenz, M.

Blood 81, 2854-2859, 1993

A:Title: Drosophila forkhead homologues are expressed in a lineage-restricted manner

A:Reference number: A48924; MUID:93271467

A:Accession: B48924

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-108 <HRO>

A:Experimental source: HEL erythroleukemia cell line

A:Note: sequence extracted from NCBI backbone (NCBIP:133164)

C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology

F:9-100/Domain: fork head DNA-binding domain homology <FHD>

Query Match

Best Local Similarity 7.1%; Score 39; DB 2; Length 108;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 PFYRDKNQGMQNSIRHNSLNECFVKVPRDDKPKGKSGY 151

Db 44 PFYRDKNQGMQNSIRHNSLNECFVKVPRDDKPKGKSGY 82

RESULT 6

E56556

fork head homolog (clone gXPD-4) - African clawed frog (fragment)

C:Species: Xenopus laevis (African clawed frog)

C>Date: 05-Jan-1996 #sequence\_revision 05-Jan-1996 #text\_change 12-Sep-1997

C:Accession: E56556

R:Knöchel, S.; Lef, J.; Clement, J.; Klocke, B.; Hille, S.; Koster, M.; Knöchel, W.

Mech. Dev. 38, 157-165, 1992

A:Title: Activin A induced expression of a fork head related gene in posterior chorda

A:Reference number: A56556; MUID:93041288

A:Accession: E56556

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-111 <KNO>

A:Note: sequence extracted from NCBI backbone (NCBIP:118181)

C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology

F:10-101/Domain: fork head DNA-binding domain homology <FHD>

Query Match

Best Local Similarity 6.3%; Score 35; DB 2; Length 111;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 NKQGMQNSIRHNSLNECFVKVPRDDKPKGKSGY 152

Db 50 NKQGMQNSIRHNSLNECFVKVPRDDKPKGKSGY 84

## RESULT 7

559870  
 fork head domain protein crocodile - fruit fly (*Drosophila melanogaster*)  
 C:Species: *Drosophila melanogaster*  
 C>Date: 19-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 21-Jul-2000  
 C:Accession: S59870; A46178  
 R:Kestner, K.H.; Lee, K.H.; Schlondorff, J.; Hiemisch, H.; Monaghan, A.P.; Schutz, G.  
 Proc. Natl. Acad. Sci. U.S.A. 90, 7628-7631, 1993  
 A:Title: Six members of the mouse forkhead gene family are developmentally regulated.  
 A:Reference number: A47746; MUID:93361500  
 A:Accession: S59870; MUID:96080166  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-508 <HAES>  
 R:Hacker, U.; Grossniklaus, U.; Gehring, W.J.; Jackle, H.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 8754-8758, 1992  
 A:Title: Developmentally regulated *Drosophila* gene family encoding the fork head domain.  
 A:Reference number: A46178; MUID:92409595  
 A:Accession: A46178  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid  
 A:Residues: 55-182 <HAG>  
 A:Cross-references: GB:96440; NID:9157425; PIDN:AAF02177.1; PID:96042185  
 A:Note: sequence extracted from NCBI backbone (NCBIF:114222)  
 C:Genetics:  
 A:Gene: *croc*  
 A:Cross-references: FlyBase:FBgn0014143  
 C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
 F:70-161/Domain: fork head DNA-binding domain homology <FHD>

Query Match 4.5%; Score 25; DB 2; Length 508;  
 Best Local Similarity 100.0%; Pred. No. 3e-13; Indels 0; Gaps 0;  
 Matches 25; Conservative 0; Mismatches 0;

OY 115 YRDNKGWNSIRHNLSECFVKV 139  
 ||||||||||||||||||||  
 Db 107 YRDNKGWNSIRHNLSECFVKV 131

## RESULT 8

447746  
 forkhead transcription factor fkh-3 - mouse (fragment)  
 C:Species: *Mus musculus* (house mouse)  
 C>Date: 03-Mar-1994 #sequence\_revision 05-Jan-1996 #text\_change 24-Sep-1999  
 C:Accession: C47746; S36072  
 R:Kestner, K.H.; Lee, K.H.; Schlondorff, J.; Hiemisch, H.; Monaghan, A.P.; Schutz, G.  
 Proc. Natl. Acad. Sci. U.S.A. 90, 7628-7631, 1993  
 A:Title: Six members of the mouse forkhead gene family are developmentally regulated.  
 A:Reference number: A47746; MUID:93361500  
 A:Accession: C47746  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-111 <KAE>  
 A:Cross-references: EMBL:X71944; NID:9311740; PIDN:CAA50746.1; PID:9311741  
 A:Note: submitted to the EMBL Data Library, May 1993  
 A:Note: sequence extracted from NCBI backbone (NCBIF:136532)  
 C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
 C:Keywords: DNA binding; transcription factor  
 F:10-101/Domain: fork head DNA-binding domain homology <FHD>

Query Match 4.2%; Score 23; DB 2; Length 111;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-12; Indels 0; Gaps 0;  
 Matches 23; Conservative 0; Mismatches 0;

OY 121 GMONSIRHNLSECFVKVPRDD 143  
 ||||||||||||||||||||  
 Db 53 GMONSIRHNLSECFVKVPRDD 75

## RESULT 9

F47746  
 forkhead transcription factor fkh-6 - mouse (fragment)  
 C:Species: *Mus musculus* (house mouse)  
 C>Date: 03-Mar-1994 #sequence\_revision 05-Jan-1996 #text\_change 24-Sep-1999  
 C:Accession: F47746; S36075  
 R:Kestner, K.H.; Lee, K.H.; Schlondorff, J.; Hiemisch, H.; Monaghan, A.P.; Schutz, G.  
 Proc. Natl. Acad. Sci. U.S.A. 90, 7628-7631, 1993  
 A:Title: Six members of the mouse forkhead gene family are developmentally regulated.  
 A:Reference number: A47746; MUID:93361500  
 A:Accession: F47746  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-111 <KAE>  
 A:Cross-references: EMBL:X71944; NID:9311746; PIDN:CAA50746.1; PID:9311747  
 A:Note: submitted to the EMBL Data Library, May 1993  
 A:Note: sequence extracted from NCBI backbone (NCBIF:136535)  
 C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
 C:Keywords: DNA binding; transcription factor  
 F:10-101/Domain: fork head DNA-binding domain homology <FHD>

Query Match 4.0%; Score 22; DB 2; Length 111;  
 Best Local Similarity 100.0%; Pred. No. 3e-11; Indels 0; Gaps 0;  
 Matches 22; Conservative 0; Mismatches 0;

OY 120 GGMNSIRHNLSECFVKVPR 141  
 ||||||||||||||||||||  
 Db 52 GGMNSIRHNLSECFVKVPR 73

## RESULT 10

T16880  
 hypothetical protein T14G12.4 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jan-2000  
 C:Accession: T16880  
 R:Willcox, L.  
 submitted to the EMBL Data Library, November 1995  
 A:Description: The sequence of *C. elegans* cosmid T14G12.  
 A:Reference number: Z18596  
 A:Accession: T16880  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-270 <WLL>  
 A:Cross-references: EMBL:U41268; NID:91086843; PID:91086847; PIDN:AA82436.1; CESP:T1  
 C:Genetics:  
 A:Gene: CESP:T14G12.4  
 A:Introns: 37/1; 72/3; 164/1  
 C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
 F:93-185/Domain: fork head DNA-binding domain homology <FHD>

Query Match 3.8%; Score 21; DB 2; Length 270;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-10; Indels 0; Gaps 0;  
 Matches 21; Conservative 0; Mismatches 0;

OY 113 PFYRDNKGWNSIRHNLSTLN 133  
 ||||||||||||||||||||  
 Db 128 PFYRDNKGWNSIRHNLSTLN 148

## RESULT 11

S23055  
 SLP2 protein - fruit fly (*Drosophila melanogaster*)  
 C:Species: *Drosophila melanogaster*  
 C>Date: 19-Feb-1994 #sequence\_revision 01-Sep-1995 #text\_change 24-Sep-1999  
 C:Accession: S23055  
 R:Grossniklaus, U.; Pearson, R.K.; Gehring, W.J.  
 Genes Dev. 6, 1030-1051, 1992  
 A:Title: The *Drosophila* sloppy paired locus encodes two proteins involved in segmenta  
 A:Reference number: S23053; MUID:92275347  
 A:Accession: S23055  
 A:Status: preliminary

A:Molecule type: DNA  
 A:Residues: 1-445 <GRO>  
 A:Cross-references: EMBL:X66098; NID:98620; PIDN:CAA46892.1; PID:98621  
 C:Genetics:  
 A:Gene: FLYBase:slp2  
 A:Cross-references: FLYBase:FBgn0004567  
 C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
 C:Keywords: DNA binding; transcription regulation  
 F:180-271/Domain: fork head DNA-binding domain homology <FHD>

Query Match  
 Best Local Similarity 3.4%; Score 19; DB 2; Length 445;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 YRDNKQGMNSIRHNSLN 133  
 Db 217 YRDNKQGMNSIRHNSLN 235

## RESULT 12

S23056

slp2 protein - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 19-Feb-1994 #sequence\_revision 01-Sep-1995 #text\_change 24-Sep-1999

C:Accession: S23056

R:Grossniklaus, U.; Pearson, R.K.; Gehring, W.J.

A:Title: The Drosophila sloppy paired locus encodes two proteins involved in segmental

A:Reference number: S23053; MUID:92275347

A:Accession: S23056

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-445 &lt;GRO&gt;

A:Cross-references: EMBL:X66097; NID:98622; PIDN:CAA46891.1; PID:98623

C:Genetics:  
 A:Gene: FLYBase:slp2  
 A:Cross-references: FLYBase:FBgn0004567  
 C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
 F:180-271/Domain: fork head DNA-binding domain homology <FHD>

Query Match  
 Best Local Similarity 3.4%; Score 19; DB 2; Length 445;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 YRDNKQGMNSIRHNSLN 133  
 Db 217 YRDNKQGMNSIRHNSLN 235

## RESULT 13

S51630

FRPAC-7 protein - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 15-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 24-Sep-1999

C:Accession: S51630

R:Pierrou, S.; Heliqvist, M.; Samuelsson, L.; Ennerbeck, S.; Carlsson, P.

A:Title: Cloning and characterization of seven human forkhead proteins: binding site spe

A:Reference number: S51630; MUID:95045392

A:Accession: S51630

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-106 &lt;PIE&gt;

A:Cross-references: EMBL:U13225; NID:9563169; PIDN:AAA92042.1; PID:9563170

A:Gene: GDB:FKHL1; FRPAC7

A:Cross-references: GDB:450231

C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
 F:6-97/Domain: fork head DNA-binding domain homology <FHD>

Query Match  
 Best Local Similarity 2.9%; Score 16; DB 2; Length 106;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 TINGYOFIMDRFPY 115  
 Db 28 TINGYOFIMDRFPY 43

## RESULT 14

B54743

transcription factor HRF2 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 12-Sep-1997

C:Accession: B54743

R:Murphy, D.B.; Wiese, S.; Burfield, P.; Schmundt, D.; Mattei, M.G.; Schulz-Schaeffer

A:Title: Human brain factor 1, a new member of the fork head gene family.

A:Reference number: A54743; MUID:95048332

A:Accession: B54743

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-109 &lt;MUR&gt;

A:Cross-references: GB:X74143

C:Genetics:  
 A:Gene: GDB:FKHL2; HRF-2; HRF2  
 A:Cross-references: GDB:375747; OMIM:600779  
 A:Map position: 14q11-14q13  
 C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
 F:8-99/Domain: fork head DNA-binding domain homology <FHD>

Query Match  
 Best Local Similarity 2.9%; Score 16; DB 2; Length 109;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 NKQGMNSIRHNSLN 133  
 Db 48 NKQGMNSIRHNSLN 63

## RESULT 15

B46178

Probable transcription factor fork head domain 2 (FD2) - fruit fly (Drosophila melano

C:Species: Drosophila melanogaster

C:Date: 22-Sep-1993 #sequence\_revision 25-Apr-1997 #text\_change 24-Sep-1999

C:Accession: B46178

R:Hacker, U.; Grossniklaus, U.; Gehring, W.J.; Jackle, H.

A:Title: Developmentally regulated Drosophila gene family encoding the fork head doma

A:Reference number: A46178; MUID:92409595

A:Accession: B46178

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-134 &lt;HAC&gt;

A:Cross-references: GB:M6441; NID:9157426; PIDN:AAA28533.1; PID:9157427

A:Note: Sequence extracted from NCBI backbone (NCBI:114224)

C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
 F:16-113/Domain: fork head DNA-binding domain homology <FHD>

Query Match  
 Best Local Similarity 2.9%; Score 16; DB 2; Length 134;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 NKQGMNSIRHNSLN 133  
 Db 56 NKQGMNSIRHNSLN 71

Wed Feb 7 14:05:59 2001

Search completed: February 6, 2001, 23:15:18  
Job time: 8142 sec

us-09-292-862-2.rpt



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 6, 2001, 22:52:59 ; Search time 49.7 Seconds  
(without alignments)  
359.329 Million cell updates/sec

Title: US-09-292-862-2

Sequence: 553  
1 MQARYSVSPNSLGVVPLG.....PSSQSLYRTSGAFVYDCKSF 553

Scoring table: OLIGO  
Gapop 60.0, Gapext 60.0

Searched: 88757 seqs, 32294092 residues

Word size: 0

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	250	45.2	553	1 FXCL_HUMAN	Q12948 homo sapien
2	120	21.7	553	1 FXCL_MOUSE	Q61572 mus musculu
3	61	11.0	494	1 FXC2_MOUSE	Q61850 mus musculu
4	61	11.0	501	1 FXC2_HUMAN	Q99958 homo sapien
5	59	10.7	101	1 FXC2_RAT	Q63246 rattus norv
6	25	4.5	508	1 CROC_DROME	P32027 drosophila
7	23	4.2	111	1 FREX_HUMAN	Q61574 mus musculu
8	23	4.2	111	1 FREX_MOUSE	Q61573 mus musculu
9	22	4.0	337	1 SLP2_DROME	P32031 drosophila
10	19	3.4	106	1 FXL1_HUMAN	Q02361 homo sapien
11	16	2.9	134	1 FD2_DROME	Q02360 drosophila
12	16	2.9	134	1 QIN_AVIS3	P56260 avian sarco
13	16	2.9	451	1 FXGB_CHICK	Q90964 gallus gall
14	16	2.9	459	1 FXGB_HUMAN	P55315 homo sapien
15	16	2.9	477	1 FXGB_HUMAN	P55315 homo sapien
16	16	2.9	480	1 FXGB_RAT	Q00939 rattus norv
17	16	2.9	481	1 FXGB_MOUSE	Q63245 mus musculu
18	16	2.9	481	1 FXGB_MOUSE	Q63245 mus musculu
19	15	2.7	101	1 FXGA_RAT	Q63251 rattus norv
20	15	2.7	101	1 FXGA_RAT	Q63251 rattus norv
21	15	2.7	128	1 FD3_DROME	Q02361 drosophila
22	15	2.7	394	1 FXGA_CHICK	P79772 gallus gall
23	15	2.7	440	1 FXD1_MOUSE	Q98937 gallus gall
24	15	2.7	440	1 FXD1_MOUSE	Q98937 gallus gall
25	15	2.7	465	1 FXD1_HUMAN	Q61345 mus musculu
26	15	2.7	465	1 FXD1_HUMAN	Q61345 mus musculu
27	15	2.7	478	1 FXD2_HUMAN	Q61345 mus musculu
28	15	2.7	497	1 FXD3_HUMAN	Q61345 mus musculu
29	14	2.5	444	1 FXD4_MOUSE	Q63248 rattus norv
30	14	2.4	444	1 FXD4_MOUSE	Q63248 rattus norv
31	13	2.4	101	1 FXL1_RAT	P32030 drosophila
32	13	2.4	351	1 FXL1_HUMAN	Q12951 homo sapien
33	13	2.4	351	1 FXL1_HUMAN	Q12951 homo sapien

34	13	2.4	443	1 OC3N_HUMAN	P20265 homo sapien
35	13	2.4	445	1 OC3N_MOUSE	P31360 mus musculu
36	13	2.4	543	1 YP91_MYCTU	Q50630 mycobacteri
37	13	2.4	617	1 FXK1_MOUSE	P42128 mus musculu
38	13	2.4	655	1 ILE1_HUMAN	Q01167 homo sapien
39	13	2.4	862	1 FXH2_YEAST	P41813 saccharomyc
40	12	2.2	106	1 FXD4_HUMAN	Q12950 homo sapien
41	12	2.2	421	1 FXD1_HUMAN	Q92949 homo sapien
42	12	2.2	421	1 FXD1_MOUSE	Q61660 mus musculu
43	12	2.2	421	1 FXD1_RAT	Q63247 rattus norv
44	12	2.2	558	1 ROL_HUMAN	P14866 homo sapien
45	12	2.2	564	1 HCM1_YEAST	P25364 saccharomyc

## ALIGNMENTS

RESULT 1  
ID FXCL\_HUMAN STANDARD: PRT: 553 AA.  
AC 012948: Q99P06:  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE FORKHEAD BOX PROTEIN C1 (FORKHEAD-RELATED PROTEIN FKHL7) (FORKHEAD-RELATED TRANSCRIPTION FACTOR 3) (FREAC-3).  
DE FOXCL OR FKHL7 OR FREAC3.  
GN Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
[1]  
RN SEQUENCE FROM N.A., AND VARIANTS S-112; M-126 AND L-131.  
RX MEDLINE-98282091; PubMed-9620769;  
RA Nishimura D.Y., Swiderski R.E., Alward W.L.M., Searby C.C.,  
RA Pettit S.R., Bennett S.R., Kanis A.B., Gastier J.M., Stone E.M.,  
RA Sheffield V.C.  
RT The forkhead transcription factor gene FKHL7 is responsible for  
RT glaucoma phenotypes which map to 6p25.  
RL Nat. Genet. 19:140-147(1998).  
[2]  
RN SEQUENCE FROM N.A., AND VARIANTS ARA THR-82 AND MET-87.  
RX MEDLINE-99011252; PubMed-9792859;  
RA Mearns A.J., Jordan T., Mirzayans F., Dubois S., Kume T., Parlee M.,  
RA Ritch R., Koop B., Kuo W.-L., Collins C., Marshall J., Gould D.B.,  
RA Pearce W., Raymond P., Enerbaeck S., Morissette J., Bhattacharya S.,  
RA Hogan B., Raymond V., Walter M.A.;  
RT Mutation of the forkhead/winged helix gene, FKHL7, in patients with  
RT Axenfeld-Rieger anomaly.  
RL Am. J. Hum. Genet. 63:1316-1328(1998).  
[3]  
RN SEQUENCE OF 73-178 FROM N.A.  
RX MEDLINE-95045392; PubMed-7957066;  
RA Pierron S., Heilly M., Samuelsson L., Enerbaeck S., Carlsson P.;  
RT Identifying and characterizing a group of seven human forkhead proteins:  
RT binding site specificity and DNA bending.  
RL EMBO J. 13:5002-5012(1994).  
[4]  
RN FUNCTION: BINDING OF FREAC-3 AND FREAC-4 TO THEIR COGNATE SITES  
RN RESULTS: BINDING OF THE DNA AT AN ANGLE OF 80-90 DEGREES.  
[5]  
RN SUBCELLULAR LOCATION: NUCLEAR.  
[6]  
RN DISEASE: DEFECTS IN FKHL7 ARE THE CAUSE A SPECTRUM OF GLAUCOMA  
RN PHENOTYPES SUCH AS AXENFELD-RIEGER ANOMALY (ARA), AXENFELD-RIEGER  
RN SYNDROME (ARS) AND IRIDOGONIODYSGENESIS ANOMALY (IGDA). ARS IS AN  
RN AUTOSOMAL DOMINANT DISORDER PRESENTING WITH ARA-LIKE OCULAR  
RN FINDINGS IN ADDITION TO ABNORMALITIES OF THE TEETH, JAW AND  
RN UMBILICUS.  
[7]  
RN SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
[8]  
RN THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
RN between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
RN the European Bioinformatics Institute. There are no restrictions on its  
RN use by non-profit institutions as long as its content is in no way  
RN modified and this statement is not removed. Usage by and for commercial

DR EMBL; AF048693; AAC18081.1; -  
DR EMBL; U13221; AAA92038.1; -  
DR EMBL; AF078096; AAC72915.1; -  
DR MIM; 601090; -

```
DR  MIM: 601631; -
DR  INTERPRO: IPR001766; -
DR  INTERPRO: IPR002953; -
DR  INTERPRO: IPR002965; -
DR  PfAM: PF00250; Fork_head, 1.
DR  PRINTS: PRO0053; ForkHEAD.
DR  PRINTS: PRO1217; PRICHHEAD.
DR  PRINTS: PRO1228; EGGSHEET.
```

DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
DR PROSITE: PS50039; FORK\_HEAD\_3; 1.  
KW DNA-binding; Nuclear protein; Transcription regulation  
KW Disease mutation.  
FT DNA BIND 77 170

Protein	Domain	Position	Residue	Frequency	Conservation	Similarity	Identity
FT	DOMAIN	28	106	100%	100%	100%	100%
FT	DOMAIN	28	33	100%	100%	100%	100%
FT	DOMAIN	169	173	100%	100%	100%	100%
FT	DOMAIN	194	197	100%	100%	100%	100%
FT	DOMAIN	262	272	100%	100%	100%	100%
FT	DOMAIN	292	297	100%	100%	100%	100%
FT	DOMAIN	375	382	100%	100%	100%	100%
FT	DOMAIN	438	445	100%	100%	100%	100%
FT	DOMAIN	447	456	100%	100%	100%	100%
FT	DOMAIN	486	495	100%	100%	100%	100%
FT	VARIANT	82	82	100%	100%	100%	100%

Sequence	535 AA	56787 MW	DB081B94.40303369
VARIANT	87	87	/FTID=VAR_007944. I -> M (IN AR).
VARIANT	112	112	/FTID=VAR_007945. E -> S (IN IGDA).
VARIANT	126	126	/FTID=VAR_007815. I -> M (IN AR).
VARIANT	131	131	/FTID=VAR_007816. S -> L (IN AR).
CONFLICT	180	180	/FTID=VAR_007817. V -> L (IN REF. 2).
CONFLICT	199	202	RODP -> ASPR (IN REF. 2)
CONFLICT	426	426	N -> D (IN REF. 2)
SEQUENCE	535 AA	56787 MW	DB081B94.40303369

Query Match	45.28;	Score 250;	DB 1;	Length 553;
Best Local Similarity	99.78;	Pred. No. 1.6e-192;		
Matches 350;	Conservative	0;	Mismatches	1

[illegible]

```

      2
EXCL_MOUSE
ID EXCL_MOUSE STANDARD: PRT; 553 AA.
AC 061572; 061582; 088409;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE FORKHEAD BOX PROTEIN C1 (FORKHEAD-RELATED PROTEIN FKHL7) (FORKHEAD-RELATED TRANSCRIPTION FACTOR 3) (FRAC-3) (TRANSCRIPTION FACTOR FKH-1) (MESODERM/MESENCHYME FORKHEAD 1) (MF-1).
GN FOXO1 OR FKHL7 OR FRAC3 OR FKH1 OR MF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98297351; PubMed=6535428;
RA Kume T., Deng K.Y., Winfrey V., Gould D.B., Walter M.A., Hogan B.L.M.;
RT "The forkhead/winged helix gene Mf1 is disrupted in the pleiotropic RT mouse mutation congenital hydrocephalus."
RL Cell 93:985-996(1998).
RN [2]
RP SEQUENCE OF 69-179 FROM N.A.
RX STRAIN=129;
RC MEDLINE=93361500; PubMed=7689224;
RA Kaestner K.H., Lee K.H., Schloendorff J., Hiemisch H.,
RA Monaghan A.P., Schuetz G.;
RT "Six members of the mouse forkhead gene family are developmentally regulated".
RL Proc. Natl. Acad. Sci. U.S.A. 90:7628-7631(1993).
RN [3]
RP SEQUENCE OF 71-187 FROM N.A.
RX MEDLINE=93387221; PubMed=8375339;
RA Sasaki H., Hogan B.L.;
RT "Differential expression of multiple fork head related genes during RT gastrulation and axial pattern formation in the mouse embryo.",
RL Development 118:47-59(1993).
CC -I SUBCELLULAR LOCATION: NUCLEAR.
CC -I TISSUE SPECIFICITY: EXPRESSED IN MANY EMBRYONIC TISSUES, INCLUDING PREHODRGENIC MESENCHYME, PERIOULAR MESENCHYME, MENINGES, ENDOTHELIAL CELLS, AND KIDNEY.
CC -I DEVELOPMENTAL STAGE: EXPRESSED DURING EMBRYOGENESIS.
CC -I SIMILARITY: CONTAINS 1 FORK HEAD DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration CC between the Swiss Institute of Bioinformatics and the EMBL outstation CC the European Bioinformatics Institute. There are no restrictions on its CC use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
DR EMBL; AF045017; AAC24209.1; --
DR EMBL; L10406; AAA03159.1; --
DR EMBL; X71939; CAAS0741.1; --
DR HSSP; O63243; ZHPH.
DR TRANSFAC; T02426; --
DR MGD; MG1:1347466; FOXC1.
DR INTERPRO; IPR001766; --
DR PFAM; PF00250; Fork_head_1.
DR PRINTS; PR00053; FORKHEAD
DR PROSITE; PS00657; FORK_HEAD_1.1
DR PROSITE; PS00658; FORK_HEAD_2.1
DR PROSITE; PS50039; FORK_HEAD_3.1.
KW DNA-binding; Nuclear protein; Transcription regulation.
FT DOMAIN 28 33 POLY-ALA
FT DNA_BIND 77 168 FORK-HEAD.
FT DOMAIN 169 173 POLY-ARG.
FT DOMAIN 194 197 POLY-PRO.
FT DOMAIN 264 274 POLY-SER.
FT DOMAIN 375 386 POLY-GLY.

```



FT DOMAIN 444 451 POLY-SER.  
 FT DOMAIN 453 456 POLY-GLY.  
 FT DOMAIN 486 496 POLY-ALA.  
 FT CONFLICT 180 187 VKDKEEG-> KKEITFIG (IN REF. 3).  
 FT SEQUENCE 553 AA: 56953 MW: 3CDD1EF9CA4F217 CRC64;

Query Match 21.7%: Score 120; DB 1; Length 553;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-88;  
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 MARAGPTTPOPKDMVKPPYSYIALITMAIONAPDKITLITGTYOFIMDFPPYRDNK 119  
 DB 60 MARAGPTTPOPKDMVKPPYSYIALITMAIONAPDKITLITGTYOFIMDFPPYRDNK 119  
 OY 120 OGOMNSIRHNLNLNCECFYKVPDRDKKPGKGSYWTLPDPSYNMFENGSLRRRRRKKKA 179  
 DB 120 OGOMNSIRHNLNLNCECFYKVPDRDKKPGKGSYWTLPDPSYNMFENGSLRRRRRKKKA 179

RESULT 3  
 FXC2\_MOUSE STANDARD: PRT: 494 AA.

ID FXC2\_MOUSE 061850; 063869; P97948;  
 AC 01-NOV-1997 (Rel. 35; Created)  
 DT 15-JUL-1998 (Rel. 36; Last sequence update)  
 DT 30-MAY-2000 (Rel. 39; Last annotation update)  
 DE FORKHEAD BOX PROTEIN C2 (FORKHEAD-RELATED PROTEIN FKHL14) (MESENCHYME  
 DE FORK HEAD PROTEIN 1) (MFH-1 PROTEIN) (TRANSCRIPTION FACTOR FRH-14).  
 GN FOXC2 OR FKHL14 OR MFH1 OR FKHL14.  
 OS Mus musculus (Mouse);  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP STRAIN=C57BL/6;  
 RC Kaestner K.H., Blackmann S., Monaghan A.P., Schlondorff J.,  
 RA Muncheva A., Lichter P., Schütz G.;  
 RA Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=129;  
 RX MEDLINE=97312712; PubMed=9169153;  
 RT Miura N., Iida K., Kakimura H., Yang X.-L., Sugiyama T.;  
 RT "Isolation of the mouse (MFH-1) and human (FKHL14) mesenchyme fork  
 RT head-1 genes reveals conservation of their gene and protein  
 RT structures";  
 RT Genomics 41:489-492(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ICR;  
 RA Miura N.;  
 RA Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 RL [4]  
 RN SEQUENCE OF 34-494 FROM N.A.  
 RX MEDLINE=93314779; PubMed=8525367;  
 RT Miura N., Wanaoka A., Tohyama M., Tanaka K.;  
 RT "MFH-1, a new member of the fork head domain family, is expressed in  
 RT developing mesenchyme";  
 RT FEBS Lett. 326:171-176(1993).  
 RL [5]  
 CC -1- FUNCTION: MIGHT BE INVOLVED IN THE FORMATION OF SPECIAL  
 CC MESENCHYMAL TISSUES.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED TEMPORALLY IN DEVELOPING EMBRYOS,  
 CC FIRST IN THE NON-NOTOCHORDAL MESODERM AND LATER IN AREAS OF  
 CC MESENCHYMAL CONDENSATION IN THE TRUNK, HEAD, AND LIMBS.  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X92499; CAA63244.1; -  
 CC EMBL: Y08222; CAA69399.1; -  
 CC EMBL: X74040; CAA52192.1; -  
 CC EMBL: S63607; AAB27463.1; -  
 CC TRANSFAC: T02445; -  
 CC MGD: MGI:1347481; FOXC2.  
 CC INTERPRO: IPR001766; -  
 CC PFAM: PF00250; Fork head; 1.  
 CC PRINTS: PR00053; FORKHEAD.  
 CC PROSITE: PS00657; FORK HEAD 1; 1.  
 CC PROSITE: PS00658; FORK HEAD 2; 1.  
 CC PROSITE: PS00039; FORK HEAD 3; 1.  
 CC Transcription regulation; DNA binding; Nuclear protein.  
 FT DNA BIND 70 161 FORK-HEAD.  
 FT DOMAIN 162 166 POLY-ARG.  
 FT DOMAIN 386 395 HIS-RICH.  
 FT DOMAIN 396 415 ALA/PRO-RICH.  
 FT CONFLICT 12 13 AL -> V (IN REF. 1).  
 FT SEQUENCE 494 AA: 52874 MW: 232CD0D6FA64320A6 CRC64;

Query Match 11.0%; Score 61; DB 1; Length 494;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-41;  
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 118 NKOGMNSIRHNLNLNCECFYKVPDRDKKPGKGSYWTLPDPSYNMFENGSLRRRRRKKK 177  
 DB 111 NKOGMNSIRHNLNLNCECFYKVPDRDKKPGKGSYWTLPDPSYNMFENGSLRRRRRKKK 170  
 OY 178 D 178  
 DB 171 D 171

RESULT 4  
 FXC2\_HUMAN STANDARD: PRT: 501 AA.

ID FXC2\_HUMAN 099958;  
 AC 15-JUL-1998 (Rel. 36; Created)  
 DT 15-JUL-1998 (Rel. 36; Last sequence update)  
 DT 30-MAY-2000 (Rel. 39; Last annotation update)  
 DE FORKHEAD BOX PROTEIN C2 (FORKHEAD-RELATED PROTEIN FKHL14) (MESENCHYME  
 DE FORK HEAD PROTEIN 1) (MFH-1 PROTEIN) (TRANSCRIPTION FACTOR FRH-14).  
 GN FOXC2 OR FKHL14 OR MFH1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97312712; PubMed=9169153;  
 RT Miura N., Iida K., Kakimura H., Yang X.-L., Sugiyama T.;  
 RT "Isolation of the mouse (MFH-1) and human (FKHL14) mesenchyme fork  
 RT head-1 genes reveals conservation of their gene and protein  
 RT structures";  
 RT Genomics 41:489-492(1997).  
 RL [2]  
 CC -1- FUNCTION: MIGHT BE INVOLVED IN THE FORMATION OF SPECIAL  
 CC MESENCHYMAL TISSUES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: Y08223; CAA69400.1; -  
 DR MIM: 602402; -

```

RESULT 5
EXFC2_RAT
ID      EXFC2_RAT      STANDARD:      PRT;      101 AA.
AC      063246;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      FORKHEAD BOX PROTEIN C2 (BRAIN FACTOR-3) (BF-3) (HFBF-BF-3) (FRAGMENT).
GN      FOXK2 OR HFBFBF.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
PN      [1]
RP      SEQUENCE FROM N.A.
RA      STRAIN=SPRAGUE-DAWLEY;
RA      Lai E., Prezioso V.R., Tao W., Chen W.S., Darnell J.E. Jr.;
RL      Submitted (OCT-1993) to the EMBL/Genbank/DDJ databases.
CC      -1- FUNCTION: MIGHT BE INVOLVED IN THE FORMATION OF SPECIAL
CC      MESONEURAL TISSUES (BY SIMILARITY).
CC      -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC      -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      at the European Bioinformatics Institute. There are no restrictions on
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: L13193; AAA11320.1; -
DR      HSSP: Q63245; 2HFH.
DR      INTERPRO: IPR001766; -
DR      PFAM: PF00250; Fork_head.1;
DR      PROSITE: PS00657; FORK_HEAD_1; 1.
DR      PROSITE: PS00658; FORK_HEAD_2; 1.
DR      PROSITE: PS50039; FORK_HEAD_3; 1.
DR      KW      Transcription regulation; DNA-binding; Nuclear protein.
FT      NON_TER      1      1
FT      DNA_BIND      2      93      FORK-HEAD.
FT      NON_TER      101      101
SQ      SEQUENCE      101 AA;      12098 MW;      7FBD5BEF53F5B01 CRC64;

```

	RESULT	6
ID	CROC.DROME	STANDARD;
AC	P32027; O9VP32;	PRT; 508 AA.
DT	01-JUL-1993	(Rel. 26, Created)
DT	01-NOV-1997	(Rel. 35, Last sequence update)
DT	01-OCT-2000	(Rel. 40, Last annotation update)
DE	FORK HEAD DOMAIN PROTEIN CROCODILE (FKH PROTEIN FDI).	
GN	CROC OR FDI78E OR FDI.	
OC	Drosophila melanogaster (fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC	pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha	
OC	Ephydroidea; Drosophilidae; Drosophila.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-CANTON-S;	
RA	MEDLINE=96080166; PubMed=7489720;	
RA	Haecker U., Kaufmann E., Hartmann C., Jurgens G., Knoechel W.,	
RA	Jaackle H.;	
RT	"The Drosophila fork head domain protein crocodile is required for	
RT	the establishment of head structures.";	
RT	EMBO J. 14:5306-5317(1995).	
RN	[2]	

RP SEQUENCE FROM N.A.  
RC STRAIN-BERKELEY.  
RX MEDLINE-20196006: PubMed-10731132.  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Mortland J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blaise R.C., Champe M., Pfeiffer B.D.,  
RA Man R.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,  
RA April J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doug L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durkin K.J., Evangelista C.C., Ferraz C., Petterla S., Fleischmann W.,  
RA Foster C., Garfield A.E., Gang N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,  
RA Jajalini M., Kalush F., Kapran G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Lin X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mishina N.V., Moberg C., Morris J., Moshfegh A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclby J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Palmett K., Remington K., Saunders R.D.C., Schaefer F., Shen H.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svaydas R., Tector C., Turner R., Venter J., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zeyvel J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
LT Science 287:2185-2195(2000).

[3]  
 RN SEQUENCE OF 55-182 FROM N.A.  
 RP MEDLINE=92409595; PubMed=1356269;  
 RX Haacker U., Grossniklaus U., Gehring W.J., Jaechle H.:  
 RA "Developmentally regulated Drosophila gene family encoding the fork  
 RT head domain";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:8754-8758(1992).  
 CC -1- FUNCTION: REQUIRED FOR THE ESTABLISHMENT OF HEAD STRUCTURES.  
 CC REQUIRED TO FUNCTION AS AN EARLY PATTERNING GENE IN THE ANTERIOR OF A  
 CC MOST BLASTODERM HEAD SEGMENT ANLAGE AND FOR THE ESTABLISHMENT OF A  
 CC SPECIFIC HEAD SKELETAL STRUCTURE THAT DERIVES FROM THE NON-  
 CC ADJACENT INTERCALARY SEGMENT AT A LATER STAGE OF EMBRYOGENESIS.  
 CC BINDS THE CONSENSUS DNA SEQUENCE 5'-(AG)TAAATTC(A)-3'.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED INITIALLY IN BOTH THE ANTERIOR AND  
 CC POSTERIOR REGIONS OF THE EMBRYO.  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: S80254; AAB35643.1; -  
 CC EMBL: AF003594; AAF51727.1; -  
 CC EMBL: M96440; AAF02177.1; -  
 CC PIR: A46178; A46178.  
 CC TRANSEAC: T02291; -  
 CC FLYBASE: FBgn0014143; CROC.  
 CC INTERPRO: IPR001766; -  
 CC PFM: PF00250; Fork\_head; 1.  
 CC PRINTS: PR00053; FORKHEAD.  
 CC PROSITE: PS00657; FORK\_HEAD\_1; 1.  
 CC PROSITE: PS00658; FORK\_HEAD\_2; 1.  
 CC PROSITE: PS50039; FORK\_HEAD\_3; 1.  
 CC DR DNA-binding: Developmental protein; Nuclear protein;  
 CC KW Transcription regulation.  
 CC FT DOMAIN 34 40 POLY-ALA.  
 CC FT DNA\_BIND 69 160 FORK-HEAD.  
 CC FT DOMAIN 161 165 POLY-ARG.  
 CC FT DOMAIN 301 304 POLY-ALA.  
 CC FT DOMAIN 377 380 POLY-ASN.  
 CC FT DOMAIN 389 403 POLY-GLY.  
 CC FT DOMAIN 461 461 POLY-ALA.  
 CC FT DOMAIN 452 473 POLY-HIS.  
 CC FT DOMAIN 122 122 L->F (IN ALLELE CROC-75-3).  
 CC FT VARIANT 453 453 A->V (IN ALLELE CROC-75-3).  
 CC FT VARIANT 453 453 A->V (IN ALLELE CROC-75-3).  
 CC SEQUENCE 508 AA; 54516 MW; 2EFEDID8F63016D6 CRC64;

Query Match 4.5%; Score 25; DB 1; Length 508;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-12;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 YRDNKGQMSIRHNLSINECFVYV 139  
 DB 107 YRDNKGQMSIRHNLSINECFVYV 131

RESULT 7  
 FREA\_HUMAN STANDARD: PRT: 111 AA.  
 AC 043638;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE FORKHEAD-RELATED TRANSCRIPTION FACTOR 10 (FREC-10) (FRAGMENT).  
 GN FKHL18 OR FREAC10.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97468148; PubMed=9325056;  
 RA Cedergang A., Betz R., Lagercrantz S., Larsson C., Hjalander M.,  
 RA Carlsson P., Enemark S.;  
 RA "Chromosome localization, sequence analysis, and expression pattern  
 RT identify FKHL 18 as a novel human forkhead gene";  
 RL Genomics 44:344-346(1997).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF042831; AAC15420.1; -  
 CC MIM: 602939; -  
 CC INTERPRO: IPR001766; -  
 CC PFM: PF00250; Fork\_head; 1.  
 CC PROSITE: PS00657; FORK\_HEAD\_1; 1.  
 CC PROSITE: PS00658; FORK\_HEAD\_2; 1.  
 CC PROSITE: PS50039; FORK\_HEAD\_3; 1.  
 CC DR Transcription regulation; DNA-binding; Nuclear protein.  
 CC KW Transcription regulation; DNA-binding; Nuclear protein.  
 CC FT DNA\_BIND 1 100 FORK-HEAD.  
 CC FT NON\_TER 9 100  
 CC FT NON\_TER 111 111  
 CC SEQUENCE 111 AA; 12904 MW; BCEAID64C84B64F3 CRC64;

Query Match 4.2%; Score 23; DB 1; Length 111;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-11;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GWMSIRHNLSINECFVYVPRD 143  
 DB 53 GWMSIRHNLSINECFVYVPRD 75

RESULT 8  
 FREA\_MOUSE STANDARD: PRT: 111 AA.  
 AC 061574;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE FORKHEAD-RELATED TRANSCRIPTION FACTOR 10 (FREC-10) (FRAGMENT)  
 DE FACTOR FKHL-3 (FRAGMENT)  
 GN FKHL18 OR FREAC10 OR FKHL3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129;  
 RX MEDLINE=93361500; PubMed=7689224;  
 RA Kaestner K.H., Lee K.H., Schloendorff J., Hiemisch H.,  
 RA Monaghan A.P., Schuetz G.;  
 RA "Six members of the mouse forkhead gene family are developmentally  
 RT regulated";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:7628-7631(1993).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING EMBRYOGENESIS.  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial  
entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X71941: CAA50743.1; -  
CC TRANSFAC: T02440; -  
DR MGD: MGI:95546; FKH3.  
DR INTERPRO: IPR001766; -  
DR PFAM: PF00250; Fork\_head; 1.  
DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
DR PROSITE: PS50039; FORK\_HEAD\_3; 1.  
KW Transcription regulation; DNA-binding; Nuclear protein.  
FT DNA\_BIND 1 100 - FORK-HEAD.  
FT NON\_TER 111 111  
SQ SEQUENCE 111 AA; 12905 MW; 22A92EC48BC05CE1 CRC64;

Query Match  
Best Local Similarity 4.2%; Score 23; DB 1; Length 111;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GWMONSTRHNLINCECFVAVPRD 143  
DB 53 GWMONSTRHNLINCECFVAVPRD 75

## RESULT 9

FX1L\_MOUSE STANDARD; PRT; 337 AA.

AC 064731; -  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE FORKHEAD PROTEIN L1 (FORKHEAD-RELATED PROTEIN FKHL1) (TRANSCRIPTION  
DE FACTOR FKH-6).  
GN FOXL1 OR FKHL1 OR FKH6.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP STRAIN: CANTON N.A.  
RC STRAIN: 129;  
RA Kuestner K.H., Bleckmann S.C., Monaghan A.P., Schlondorff J.,  
RA Mincheva A., Licher P., Schuetz G.;  
RT "Clustered arrangement of winged helix genes fkh-6 and MFR-1:  
RT possible implications for mesoderm development.";  
RL Development 122:1751-1758(1996).  
RN [2]  
RP SEQUENCE OF 40-150 FROM N.A.  
RC STRAIN: 129;  
RA Kuestner K.H., Lee K.H., Schlondorff J., Hiemisch H., Monaghan A.P.,  
RA Schuetz G.;  
RT "Six members of the mouse forkhead gene family are developmentally  
RT regulated.";  
RL Proc. Natl. Acad. Sci. U.S.A. 90:7628-7631(1993).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X92498; CAA63243.1; -  
DR EMBL: X71944; CAA50746.1; -  
DR TRANSFAC: T02417; -

DR MGD: MGI:1347469; FOXL1.  
DR INTERPRO: IPR001766; -  
DR PFAM: PF00250; Fork\_head; 1.  
DR PRINTS: PR00053; FORKHEAD.  
DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
DR PROSITE: PS50039; FORK\_HEAD\_3; 1.  
KW DNA-binding; Nuclear protein; Transcription regulation.  
FT DNA\_BIND 48 139  
SQ SEQUENCE 337 AA; 36541 MW; 82FBD018B81DD6BD CRC64;

Query Match  
Best Local Similarity 4.0%; Score 22; DB 1; Length 337;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 GWMONSTRHNLINCECFVAVPR 141  
DB 91 GWMONSTRHNLINCECFVAVPR 112

## RESULT 10

SLP2\_DROME STANDARD; PRT; 445 AA.

AC E32031; -  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last annotation update)  
DE FORK HEAD DOMAIN TRANSCRIPTION FACTOR SLP2 (SLOPPY PAIRED LOCUS  
DE PROTEIN 2).  
GN SLP2.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN: CANTON-S, AND OREGON-R;  
RA MEDLINE: 92275347; PubMed-1317319;  
RA Grossniklaus U., Pearson R.K., Gehring M.J.;  
RT "The Drosophila sloppy paired locus encodes two proteins involved in  
RT segmentation that show homology to mammalian transcription factors.";  
RL Genes Dev. 6:1030-1051(1992).  
CC -1- FUNCTION: TRANSCRIPTION FACTOR INVOLVED IN SEGMENTATION. MAY  
CC OF SLP ACTIVITY SEEM TO BE REQUIRED IN DIFFERENT LEVELS  
CC OF SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE POSTERIOR HALF OF EACH  
CC PARASEGMENT JUST ANTERIOR TO THE PARASEGMENTAL BOUNDARY.  
CC -1- DEVELOPMENTAL STAGE: EXPRESSION AT 3-6 HRS OF EMBRYOGENESIS.  
CC -1- STRONG RE-EXPRESSION IN FIRST-INSTAR LARVAE.  
CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X66098; CAA46892.1; -  
DR EMBL: X66097; CAA46891.1; -  
DR PIR: S23055; S23055.  
DR TRANSFAC: T01056; -  
DR FLYBASE: FBgn0004567; slp2.  
DR INTERPRO: IPR001766; -  
DR PFAM: PF00250; Fork\_head; 1.  
DR PRINTS: PR00053; FORKHEAD.  
DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
DR PROSITE: PS50039; FORK\_HEAD\_3; 1.  
KW DNA-binding; Developmental protein; Nuclear protein;

Transcription regulation; Segmentation polarity protein.  
 KW DNAS\_BIND 179 270 FORK-HEAD.  
 FT DNAS\_BIND 94 104 ASN-RICH.  
 FT DOMAIN 362 362 R -> P (IN OREGON-R).  
 FT VARIANT 378 378 S -> N (IN OREGON-R).  
 FT VARIANT 387 387 O -> R (IN OREGON-R).  
 FT VARIANT 410 410 H -> Q (IN OREGON-R).  
 FT VARIANT 416 416 T -> I (IN OREGON-R).  
 FT VARIANT 416 416 T -> I (IN OREGON-R).  
 SQ SEQUENCE 445 AA: 50648 MW: 2CD253B7135B58D CRC64:

Query Match 3.4%; Score 19; DB 1; Length 445;  
 Best Local Similarity 100.0%; Pred. No. 7.4e-08; Indels 0; Gaps 0;  
 Matches 19; Conservative 0; Mismatches 0;

OY 115 YRDNKQWNSIRHNSLN 133  
 DB 217 YRDNKQWNSIRHNSLN 235

RESULT 11  
 FX1L\_HUMAN STANDARD; PRT; 106 AA.  
 ID 012952;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE FORKHEAD PROTEIN L1 (FORKHEAD-RELATED PROTEIN FKHL1) (FORKHEAD-RELATED TRANSCRIPTION FACTOR 7) (FREAC-7) (FRAGMENT).  
 DE RELATED TRANSCRIPTION FACTOR 7 (FREAC-7) (FRAGMENT).  
 GN FOXL1 OR FKHL1 OR FREAC7.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 CC [1]  
 RP MEDLINE=95045392; PubMed=7957066;  
 RA Pierson S., Hellqvist M., Samuelsson L., Emerbaeck S., Carlsson P.;  
 RT "Cloning and characterization of seven human forkhead proteins:  
 binding site specificity and DNA bending.";  
 RT EMBO J. 13:5002-5012(1994).  
 RL -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial/  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL: U13225; AAA92042.1;  
 DR MIM: 603252;  
 DR INTERPRO: IPR001766;  
 DR PFAM: PF00250; Fork\_head\_1; 1.  
 DR PROSITE: PS00657; FORK-HEAD\_2; 1.  
 DR PROSITE: PS00658; FORK-HEAD\_3; 1.  
 DR PROSITE: PS00039; FORK-HEAD\_3; 1.  
 DR DNAS\_BIND 1 1  
 FT DNAS\_BIND 5 96 FORK-HEAD.  
 FT NON\_TER 106 106  
 FT NON\_TER 106 106  
 SQ SEQUENCE 106 AA: 12499 MW: 22136293EBB2BF17 CRC64:

Query Match 2.9%; Score 16; DB 1; Length 106;  
 Best Local Similarity 100.0%; Pred. No. 5.8e-06; Indels 0; Gaps 0;  
 Matches 16; Conservative 0; Mismatches 0;

OY 100 TLNGIYQIMDRPPY 115  
 DB 28 TLNGIYQIMDRPPY 43

RESULT 12  
 FD2\_DROME STANDARD; PRT; 134 AA.  
 ID 002360;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 01-FEB-1995 (Rel. 31, Last annotation update)  
 DE FORK HEAD DOMAIN PROTEIN FD2 (FRAGMENT).  
 DE FD64 OR FD2.  
 OS Drosophila melanogaster (Fruit fly).  
 OS Drosophila melanogaster (Fruit fly).  
 OS Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 CC [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=92409595; PubMed=1356269;  
 RA Haacker U., Grossniklaus U., Gehring W.J., Jaekle H.;  
 RT "Developmentally regulated Drosophila gene family encoding the fork  
 RT head domain.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 89:8754-8758(1992)  
 RL -1- FUNCTION: INVOLVED IN DEVELOPMENT DURING EMBRYOGENESIS.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial/  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL: M96441; AAA28533.1;  
 DR PIR: B46178; B46178.  
 DR EMBASE: F860004895; fd64A.  
 DR INTERPRO: IPR001766;  
 DR PFAM: PF00250; Fork\_head\_1;  
 DR PROSITE: PS00657; FORK-HEAD\_1; 1.  
 DR PROSITE: PS00658; FORK-HEAD\_2; 1.  
 DR PROSITE: PS00039; FORK-HEAD\_3; 1.  
 DR DNAS\_BIND 1 1  
 FT DNAS\_BIND 15 106 FORK-HEAD.  
 FT NON\_TER 134 134  
 FT NON\_TER 134 134  
 SQ SEQUENCE 134 AA: 15798 MW: 4DFD5DE0967E308 CRC64:

Query Match 2.9%; Score 16; DB 1; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-06; Indels 0; Gaps 0;  
 Matches 16; Conservative 0; Mismatches 0;

OY 118 NKQWNSIRHNSLN 133  
 DB 56 NKQWNSIRHNSLN 71

RESULT 13  
 QIN\_AVIS3 STANDARD; PRT; 387 AA.  
 ID 012952;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE TRANSFORMING PROTEIN QIN (ONCOGENE QIN).  
 DE V-QIN.  
 OS Avian sarcoma virus (strain 31) (ASV31).  
 OS Avian sarcoma virus (strain 31) (ASV31).  
 OS Viruses; Retroviruses; Retroviridae; Avian type C retroviruses.  
 CC [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=93281605; PubMed=8099441;  
 RA Li J., Vogt P.K.;

RT "The retroviral oncogene gln belongs to the transcription factor  
 RT family that includes the homeotic gene fork head";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:4490-4494(1993).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-QIN  
 CC POLYPEPTIDE.  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: L10719; NOT ANNOTATED\_CDS.  
 DR INTERPRO: IPR001766; -  
 DR PFAM: PF00250; Fork\_head: 1.  
 DR PRINTS: PR00053; FORKHEAD.  
 DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE: PS00039; FORK\_HEAD\_3; 1.  
 KW DNA-binding; Nuclear protein; Transforming protein; Oncogene.  
 FT DOMAIN 42 54 POLY-HIS.  
 FT DOMAIN 55 58 POLY-PRO.  
 FT DOMAIN 64 67 POLY-ALA.  
 FT DOMAIN 101 106 POLY-ALA.  
 FT DNA\_BIND 142 233 FORK-HEAD.  
 SQ SEQUENCE 387 AA; 42283 MW; FEA902F50FEE42P9 CRC64;

Query Match  
 Best Local Similarity 2.9%; Score 16; DB 1; Length 387;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 NKGWONSIRHNLSTLN 133  
 DB 183 NKGWONSIRHNLSTLN 198  
 ||||||||||||||||  
 ||||||||||||||||

RESULT 14  
 EXGB\_CHICK STANDARD; PRT; 451 AA.  
 AC Q90964;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DE 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE FORKHEAD PROTEIN G1B (FORKHEAD-RELATED PROTEIN FKHL1) (TRANSCRIPTION  
 DE FACTOR BF-1) (BRAIN FACTOR 1) (BF1) (CBF-1) (PROTO-ONCOGENE C-QIN) (N-  
 DE 62-5) (CEO 3-1).  
 GN FOXG1B OR FKHL1 OR QIN.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95132616; PubMed=7831308;  
 RA Chang H.W., Li J., Kretschmar D., Vogt P.K.;  
 RT "Avian cellular homolog of the gln oncogene";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:447-451(1995).  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WHITE. LESHORN; TISSUE=RETINA;  
 RX MEDLINE=96382826; PubMed=8757134;  
 RA Yusa J., Hirano S., Yamagata M., Noda M.;  
 RT "Visual projection map specified by topographic expression of  
 RT transcription factors in the retina";  
 RL Nature 382:632-635(1996).  
 CC -1- FUNCTION: MAY DETERMINE THE NASOTEMPORAL AXIS OF THE RETINA, AND  
 CC CONSEQUENTLY SPECIFY THE TOPOGRAPHICAL PROJECTION OF THE RETINAL  
 CC GANGLION-CELL AXONS TO THE TECTUM BY CONTROLLING EXPRESSION OF

CC THEIR TARGET GENES.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 CC -1- TISSUE SPECIFICITY: RETINA AND BRAIN.  
 CC -1- DEVELOPMENTAL STAGE: CAN BE DETECTED IN REGIONS INCLUDING  
 CC PRIMORDIAL RETINA AND NEUROEPITHELIUM BY EMBRYONIC DAY 2 (E2). AT  
 CC E3, EXPRESSED IN THE NASAL RETINA AND PIGMENT EPITHELIUM AS WELL  
 CC AS IN THE TELENCEPHALON, AND AT E7 IS EXPRESSED IN RETINAL  
 CC GANGLION CELLS. LEVELS BEGIN TO DECLINE FROM E4 AND ALMOST  
 CC DISAPPEAR BY E10.  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: L36814; AAA66954.1; -  
 DR EMBL: U47275; AAB08466.1; -  
 DR INTERPRO: IPR001766; -  
 DR PFAM: PF00250; Fork\_head: 1.  
 DR PRINTS: PR00053; FORKHEAD.  
 DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE: PS00039; FORK\_HEAD\_3; 1.  
 KW Transcription regulation; DNA-binding; Nuclear protein;  
 KW developmental protein; Proto-oncogene.  
 FT DOMAIN 45 54 POLY-HIS.  
 FT DOMAIN 55 58 POLY-PRO.  
 FT DOMAIN 64 67 POLY-ALA.  
 FT DOMAIN 103 106 POLY-ALA.  
 FT DNA\_BIND 142 233 FORK-HEAD.  
 SQ SEQUENCE 451 AA; 48856 MW; E9E5B407D2321B50 CRC64;

Query Match  
 Best Local Similarity 2.9%; Score 16; DB 1; Length 451;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 NKGWONSIRHNLSTLN 133  
 DB 183 NKGWONSIRHNLSTLN 198  
 ||||||||||||||||  
 ||||||||||||||||

RESULT 15  
 EXGA\_HUMAN STANDARD; PRT; 469 AA.  
 AC P55316;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DE 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE FORKHEAD PROTEIN G1A (FORKHEAD-RELATED PROTEIN FKHL2) (TRANSCRIPTION  
 DE FACTOR BF-2) (BRAIN FACTOR 2) (BF2) (HFK2).  
 GN FOXG1A OR FKHL2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RX MEDLINE=95048332; PubMed=7959731;  
 RA Murphy D.B., Wiess S., Burfeind P., Schmundt D., Mettel M.-G.,  
 RA Schulz-Schaeffer W., Thies U.;  
 RT "Human brain factor 1, a new member of the fork head gene family";  
 RL Genomics 21:551-557(1994).  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95322450; PubMed=7599184;  
 RA Wiess S., Murphy D.B., Schilling A., Burfeind P., Schmundt D.,  
 RA Schunle V., Mettel M.-G., Thies U.;  
 RT "The genes for human brain factor 1 and 2, members of the fork head

```

RT gene family, are clustered on chromosome 14q."
RL Biochim. Biophys. Acta 1262:105-112(1995).
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ESTABLISHMENT OF THE
CC REGIONAL SUBDIVISION OF THE DEVELOPING BRAIN AND IN THE
CC DEVELOPMENT OF THE TELENCEPHALON. SEQUENCE-SPECIFIC DNA-BINDING
CC PROTEIN WITH A DISTINCT BINDING SPECIFICITY (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X78202; CAA55038.1; -
DR EMBL: X74143; CAA52240.1; -
DR TRANSFAC: T02292; -
DR MIM: 600779; -
DR INTERPRO: IPR001766; -
DR PFAM: PF00250; FORK_HEAD_1.
DR PRINTS: PR00053; FORKHEAD.
DR PROSITE: PS00657; FORK_HEAD_1; 1.
DR PROSITE: PS00658; FORK_HEAD_2; 1.
DR PROSITE: PS50039; FORK_HEAD_3; 1.
DR Transcription regulation; DNA-binding; Nuclear protein; *
KW Developmental protein.
KW
FT DOMAIN 33 56 HIS-RICH.
FT DOMAIN 57 79 PRO-RICH.
FT DOMAIN 70 73 POLY-GLN.
FT DOMAIN 83 88 POLY-ARG.
FT DNA_BIND 161 252 FORK-HEAD.
SQ SEQUENCE 469 AA; 50539 MW; 943B8BDB90008EDC CRC64;

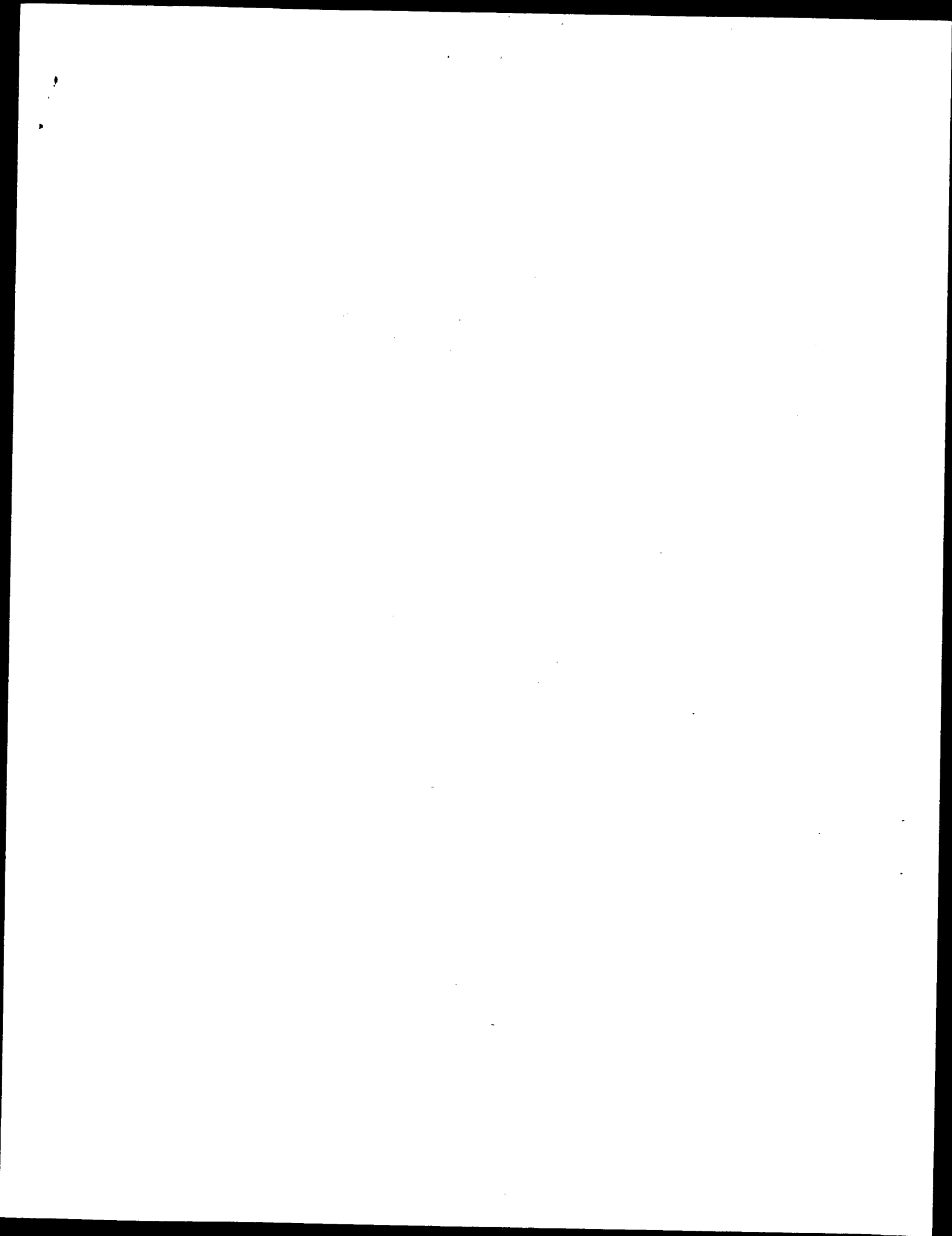
```

```

Query Match          2.9%; Score 16; DB 1; Length 469;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 118 NKCGWNSIRHNLSLN 133
   |||||
Db 202 NKCGWNSIRHNLSLN 217

```

Search completed: February 7, 2001, 00:41:05  
 Job time: 6486 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 6, 2001, 20:25:51 : Search time 51.67 Seconds  
(without alignments)  
365.960 Million cell updates/sec

Title: US-09-292-862-2

Perfect score: 553

Sequence: 1 MQARYSSSPNSLGVVPYLGL.....PSSQSLYRTSGAFVYDCKF 553

Scoring table:

OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 268485 seqs, 34193795 residues

Word size : 0

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

A.GeneSeq-36.\*  
1: /SID56/gcgdata/geneSeq/geneSeq/AA1980.DAT:\*  
2: /SID56/gcgdata/geneSeq/geneSeq/AA1981.DAT:\*  
3: /SID56/gcgdata/geneSeq/geneSeq/AA1982.DAT:\*  
4: /SID56/gcgdata/geneSeq/geneSeq/AA1983.DAT:\*  
5: /SID56/gcgdata/geneSeq/geneSeq/AA1984.DAT:\*  
6: /SID56/gcgdata/geneSeq/geneSeq/AA1985.DAT:\*  
7: /SID56/gcgdata/geneSeq/geneSeq/AA1986.DAT:\*  
8: /SID56/gcgdata/geneSeq/geneSeq/AA1987.DAT:\*  
9: /SID56/gcgdata/geneSeq/geneSeq/AA1988.DAT:\*  
10: /SID56/gcgdata/geneSeq/geneSeq/AA1989.DAT:\*  
11: /SID56/gcgdata/geneSeq/geneSeq/AA1990.DAT:\*  
12: /SID56/gcgdata/geneSeq/geneSeq/AA1991.DAT:\*  
13: /SID56/gcgdata/geneSeq/geneSeq/AA1992.DAT:\*  
14: /SID56/gcgdata/geneSeq/geneSeq/AA1993.DAT:\*  
15: /SID56/gcgdata/geneSeq/geneSeq/AA1994.DAT:\*  
16: /SID56/gcgdata/geneSeq/geneSeq/AA1995.DAT:\*  
17: /SID56/gcgdata/geneSeq/geneSeq/AA1996.DAT:\*  
18: /SID56/gcgdata/geneSeq/geneSeq/AA1997.DAT:\*  
19: /SID56/gcgdata/geneSeq/geneSeq/AA1998.DAT:\*  
20: /SID56/gcgdata/geneSeq/geneSeq/AA1999.DAT:\*  
21: /SID56/gcgdata/geneSeq/geneSeq/AA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	553	100.0	553	20	Human FREAC3 prote
2	250	45.2	553	20	Human forkhead tra
3	223	40.3	553	20	Human FKHL7 protei
4	61	11.0	501	20	FREAC11 protein
5	48	8.7	106	20	FKHL7 forkhead dom
6	40	7.2	106	20	FKHL14 forkhead do
7	17	3.1	106	20	FKHL18 forkhead do
8	16	2.9	105	20	FKHL1 forkhead do
9	16	2.9	106	20	FKHL1 forkhead do
10	16	2.9	106	20	FKHL1 forkhead do
11	16	2.9	480	14	Brain factor-1, R
12	13	2.4	76	20	FKHL10 forkhead do

13	13	2.4	106	20	V55729
14	13	2.4	540	17	W03762
15	12	2.2	106	20	V55730
16	12	2.2	106	20	V55731
17	12	2.2	106	20	V55732
18	11	2.0	37	19	W53373
19	11	2.0	37	19	W53372
20	11	2.0	37	21	V59098
21	11	2.0	531	20	V40098
22	11	2.0	595	12	R14309
23	11	2.0	595	19	W53347
24	11	2.0	595	19	V59071
25	11	2.0	921	21	V77293
26	11	2.0	921	21	V78845
27	11	2.0	4472	21	W22601
28	11	2.0	4472	18	W22611
29	11	2.0	4545	18	W67765
30	10	1.8	14	20	W92610
31	10	1.8	15	20	W67766
32	10	1.8	15	20	W32212
33	10	1.8	18	18	W32212
34	10	1.8	19	21	W83221
35	10	1.8	23	20	W67767
36	10	1.8	24	19	W79139
37	10	1.8	35	16	R80211
38	10	1.8	50	18	W13632
39	10	1.8	50	18	W13633
40	10	1.8	92	21	V44712
41	10	1.8	106	20	V55727
42	10	1.8	106	20	V55728
43	10	1.8	106	20	V55732
44	10	1.8	106	20	V55733
45	10	1.8	119	16	R99058

#### ALIGNMENTS

FKHL9 forkhead dom  
Interleukin bindin  
FKHL8 forkhead dom  
FKHL17 forkhead do  
FKHL13 forkhead do  
Nephila clavipes s  
Nephila clavipes s  
Nephila clavipes s  
N. clavipes spider  
N. clavipes spider  
Spider silk protei  
N.clavipes draglin  
Nephila clavipes s  
N. clavipes spider  
Streptomyces fradi  
KSq Arg loading di  
Tyactone synthase  
Hybrid srm6/LyG O  
Platelet glycoprot  
Mouse beta-actin a  
Platelet glycoprot  
Peptide derived fr  
Peptide linker use  
Platelet glycoprot  
G624 Gly-ala inser  
Major ampullate sp  
Apolliprotein (a)  
Shorthorn sculpin  
FKHL12 forkhead do  
FKHL5 forkhead do  
FKHL6 forkhead dom  
Spider dragline va

RESULT 1  
ID V41277 standard; Protein: 553 AA.  
AC V41277;  
XX 31-JAN-2000 (first entry)  
DT Human FREAC3 protein.  
XX  
DE  
XX  
XX FREAC3 gene; eye disease; eye developmental defect; therapy: mutation;  
KW glaucoma; anterior segment dysgenesis.  
XX  
XX Homo sapiens.  
OS  
XX  
XX  
XX M09954493-A2.  
XX  
XX 28-OCT-1999.  
XX  
XX  
XX 16-APR-1999; 99WO-IB01024.  
XX  
XX  
XX 17-APR-1998; 98US-0082206.  
XX 08-MAY-1998; 98US-0084784.  
XX  
XX (UYAL-) UNIV ALBERTA.  
XX  
XX Walter MA, Jordan T, Raymond V;  
XX  
XX MPI: 1999-634007/54.  
XX N-PSDB; 22772.  
XX  
XX Use of mutations in the FREAC3 gene for detecting predisposition to eye  
XX diseases or developmental defects and for developing products for  
XX prevention and therapy

VA3250.

Y43260;

KM lipid metabolism; adipocyte differentiation; obesity-related condition;  
KM non-insulin dependent diabetes mellitus; cardiovascular disease;  
KM catabolic condition; anorexia; bulimia; therapy.  
XX  
OS Homo sapiens.  
PN MO9854216-A1.  
XX  
PD 03-DEC-1998.  
XX  
PF 26-MAY-1998; 98MO-SE00989.  
XX  
PR 26-MAY-1997; 97SE-0001963.  
XX  
PA (PMAA ) PHARMACIA & UPJOHN AB.  
XX  
PI Carlsson P, Enerbaeck S;  
XX  
DR WPI: 1999-070211/06.  
XX  
N-PSDB: X28103.  
XX  
PT New transcription factor designated FREAC11 - which regulates  
PT adipose tissue expressed genes involved in lipid metabolism and  
PT adipocyte differentiation, used to, e.g. inhibit gene expression  
PT  
PS Claim 5; Fig 1; 23pp: English.  
XX  
SS This sequence represents the FREAC11 protein of the invention. The  
CC protein has transcriptional regulatory function directed against adipose  
CC tissue expressed genes. The protein is involved in lipid metabolism  
CC and/or adipocyte differentiation. FREAC11 is selectively expressed in  
CC adipose tissue, and is used for transcriptional regulation of adipocyte  
CC expressed genes. The DNA or protein can be used as drugs for treating  
CC obesity-related conditions to increase or decrease the activity of  
CC adipocyte expressed genes. e.g. FREAC11 can be used as a drug that blocks  
CC the cis elements of adipose expressed genes to inhibit gene expression.  
CC antisense constructs of FREAC11 can be used to down regulate expression  
CC of the FREAC11 protein can also be used for high throughput screening for  
CC substances that affect the activity of FREAC11, such as inhibitors,  
CC antagonists, or agonists. Conditions which can be treated include  
CC obesity, non-insulin dependent diabetes mellitus, cardiovascular  
CC diseases, catabolic conditions, anorexia, bulimia.  
XX  
SQ Sequence 501 AA:  
  
Query Match 11.0%; Score 61; DB 20; Length 501;  
Best Local Similarity 100.0%; Pred. NO. 6e-45;  
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 118 NKGGMNSTHNNLSLNECFVKKVPDDKRGKSGSYTLDPDSYNMFENGSPLRRRRPPKKK 177  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 112 nkggwnstihnlslnectvkvprddkpgkksywtclpdpsynmfengsfllrrrtfkxx 171  
  
QY 178 D 178  
|  
DB 172 d 172  
  
RESULT 5  
Y55723  
ID Y55723 standard; Protein: 106 AA.  
XX  
AC Y55723;  
XX  
DT 22-FEB-2000 (first entry)  
XX  
DE FKHL7 forkhead domain fragment.  
XX  
KW Forkhead transcription factor gene; FKHL7; treatment; glaucoma; human;  
KM transgenic animal; drug screening.  
XX  
OS Homo sapiens.

XX WO9953060-A2.  
 PN 21-OCT-1999.  
 XX  
 XX 14-APR-1999; 99WO-US08148.  
 PF 15-APR-1998; 98US-0081870.  
 PR 22-MAY-1998; 98US-0083352.  
 XX (IOWA ) UNIV IOWA RES FOUND.  
 PA  
 XX Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;  
 PI WPI; 1999-620429/53.  
 DR  
 XX New isolated human forkhead transcription factor gene, FKHL7, used to,  
 PT e.g. develop products for the diagnosis -  
 XX Disclosure; Fig 2; 99pp; English.  
 PS  
 CC The invention provides a human forkhead transcription factor gene, FKHL7.  
 CC The FKHL7 protein can be produced by standard recombinant methodology.  
 CC The products can be used for diagnosis, prognosis, monitoring, prevention  
 CC and treatment of glaucoma. They can also be used for the production of  
 CC transgenic animals and drug screening. Sequences Y55722-739 represent  
 CC forkhead domains of different members of the FKHL-family of genes.  
 CC  
 XX Sequence 106 AA;  
 SQ

Query Match  
 Best Local Similarity 8.7%; Score 48; DB 20; Length 106;  
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 AIONAPPKKTILNGIYQFIMDRPEFFYRDNKGQMSIRHNLINCEFY 137  
 Db 18 aignapdkkittlingiyqfimdriyrdnkqgwsirhnlisnecfv 65

RESULT 6  
 ID Y55724 standard; Protein: 106 AA.  
 AC Y55724;  
 XX  
 DT 22-FEB-2000 (first entry)  
 XX  
 DE FKHL14 forkhead domain fragment..  
 XX  
 KW Forkhead transcription factor gene; FKHL7; treatment; glaucoma; human;  
 KW transgenic animal; drug screening.  
 XX  
 OS Unidentified.  
 OS  
 PN WO9953060-A2.  
 XX 21-OCT-1999.  
 PD  
 XX 14-APR-1999; 99WO-US08148.  
 PF 15-APR-1998; 98US-0081870.  
 PR 22-MAY-1998; 98US-0083352.  
 XX (IOWA ) UNIV IOWA RES FOUND.  
 PA  
 XX Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;  
 PI WPI; 1999-620429/53.  
 DR  
 XX New isolated human forkhead transcription factor gene, FKHL7, used to,  
 PT e.g. develop products for the diagnosis -  
 XX

PS Disclosure; Fig 2; 99pp; English.  
 XX  
 CC The invention provides a human forkhead transcription factor gene, FKHL7.  
 CC The FKHL7 protein can be produced by standard recombinant methodology.  
 CC The products can be used for diagnosis, prognosis, monitoring, prevention  
 CC and treatment of glaucoma. They can also be used for the production of  
 CC transgenic animals and drug screening. Sequences Y55722-739 represent  
 CC forkhead domains of different members of the FKHL-family of genes.  
 CC  
 XX Sequence 106 AA;  
 SQ

Query Match  
 Best Local Similarity 7.2%; Score 40; DB 20; Length 106;  
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 VPRDDKKPKGKSYWTLPDPDPSYMFENGSLRRRRRFRKKD 178  
 Db 67 vprddkkpkygswtldpdpdpsymfengslrrrrfrkkkd 106

RESULT 7  
 ID Y55725 standard; Protein: 106 AA.  
 AC Y55725;  
 XX  
 DT 22-FEB-2000 (first entry)  
 XX  
 DE FKHL18 forkhead domain fragment.  
 XX  
 KW Forkhead transcription factor gene; FKHL7; treatment; glaucoma; human;  
 KW transgenic animal; drug screening.  
 XX  
 OS Unidentified.  
 OS  
 PN WO9953060-A2.  
 XX 21-OCT-1999.  
 PD  
 XX 14-APR-1999; 99WO-US08148.  
 PF 15-APR-1998; 98US-0081870.  
 PR 22-MAY-1998; 98US-0083352.  
 XX (IOWA ) UNIV IOWA RES FOUND.  
 PA  
 XX Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;  
 PI WPI; 1999-620429/53.  
 DR  
 XX New isolated human forkhead transcription factor gene, FKHL7, used to,  
 PT e.g. develop products for the diagnosis -  
 XX Disclosure; Fig 2; 99pp; English.  
 PS  
 CC The invention provides a human forkhead transcription factor gene, FKHL7.  
 CC The FKHL7 protein can be produced by standard recombinant methodology.  
 CC The products can be used for diagnosis, prognosis, monitoring, prevention  
 CC and treatment of glaucoma. They can also be used for the production of  
 CC transgenic animals and drug screening. Sequences Y55722-739 represent  
 CC forkhead domains of different members of the FKHL-family of genes.  
 CC  
 XX Sequence 106 AA;  
 SQ

Query Match  
 Best Local Similarity 3.1%; Score 17; DB 20; Length 106;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GWQMSIRHNLINCEFY 137  
 Db 49 gwqmsirhnlisnecfv 65

RESULT 8  
 Y55735 standard; Protein; 105 AA.  
 Y55735;  
 22-FEB-2000 (first entry)  
 FKHL2 forkhead domain fragment.  
 FKHL2 forkhead transcription factor gene; FKHL7; treatment; glaucoma; human;  
 transgenic animal; drug screening.  
 Unidentified.  
 WO953060-A2.  
 21-OCT-1999.  
 14-APR-1999; 99WO-US08148.  
 15-APR-1998; 98US-0081870.  
 22-MAY-1998; 98US-0083352.  
 (IOWA ) UNIV IOWA RES FOUND.  
 Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;  
 WPI; 1999-620429/53.  
 New isolated human forkhead transcription factor gene, FKHL7, used to,  
 e.g. develop products for the diagnosis -  
 Disclosure; Fig 2; 99pp; English.  
 The invention provides a human forkhead transcription factor gene, FKHL7.  
 The FKHL7 protein can be produced by standard recombinant methodology.  
 The products can be used for diagnosis, prognosis, monitoring, prevention  
 and treatment of glaucoma. They can also be used for the production of  
 transgenic animals and drug screening. Sequences Y55722-739 represent  
 forkhead domains of different members of the FKHL-family of genes.  
 Sequence 105 AA;  
 Query Match 2.9%; Score 16; DB 20; Length 105;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 118 NKQSGMNSIRHNLN 133  
 DB 46 nkqsgmnsirhnlsln 61  
 RESULT 9  
 Y55726  
 Y55726 standard; Protein; 106 AA.  
 Y55726;  
 22-FEB-2000 (first entry)  
 FKHL1 forkhead domain fragment.  
 FKHL1 forkhead transcription factor gene; FKHL7; treatment; glaucoma; human;  
 transgenic animal; drug screening.  
 Unidentified.  
 WO953060-A2.

PD 21-OCT-1999.  
 XX 14-APR-1999; 99WO-US08148.  
 XX 15-APR-1998; 98US-0081870.  
 XX 22-MAY-1998; 98US-0083352.  
 XX (IOWA ) UNIV IOWA RES FOUND.  
 XX Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;  
 XX WPI; 1999-620429/53.  
 DR New isolated human forkhead transcription factor gene, FKHL7, used to,  
 XX e.g. develop products for the diagnosis -  
 XX Disclosure; Fig 2; 99pp; English.  
 XX The invention provides a human forkhead transcription factor gene, FKHL7.  
 CC The FKHL7 protein can be produced by standard recombinant methodology.  
 CC The products can be used for diagnosis, prognosis, monitoring, prevention  
 CC and treatment of glaucoma. They can also be used for the production of  
 CC transgenic animals and drug screening. Sequences Y55722-739 represent  
 CC forkhead domains of different members of the FKHL-family of genes.  
 CC Sequence 106 AA;  
 SQ

Query Match 2.9%; Score 16; DB 20; Length 106;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 100 TLNGIYOFIMDRPPY 115  
 DB 28 tlngiyofimdrpfy 43

RESULT 10  
 Y55734  
 Y55734 standard; Protein; 106 AA.  
 Y55734;  
 22-FEB-2000 (first entry)  
 FKHL4 forkhead domain fragment.  
 FKHL4 forkhead transcription factor gene; FKHL7; treatment; glaucoma; human;  
 transgenic animal; drug screening.  
 Unidentified.  
 WO953060-A2.  
 21-OCT-1999.  
 14-APR-1999; 99WO-US08148.  
 15-APR-1998; 98US-0081870.  
 22-MAY-1998; 98US-0083352.  
 (IOWA ) UNIV IOWA RES FOUND.  
 Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;  
 WPI; 1999-620429/53.  
 New isolated human forkhead transcription factor gene, FKHL7, used to,  
 e.g. develop products for the diagnosis -  
 Disclosure; Fig 2; 99pp; English.  
 The invention provides a human forkhead transcription factor gene, FKHL7.

CC The FKHL7 protein can be produced by standard recombinant methodology.  
 CC The products can be used for diagnosis, prognosis, monitoring, prevention  
 CC and treatment of glaucoma. They can also be used for the production of  
 CC transgenic animals and drug screening. Sequences Y55722-739 represent  
 CC forkhead domains of different members of the FKHL-family of genes.  
 XX  
 SO Sequence 106 AA;

Query Match 2.9%; Score 16; DB 20; Length 106;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 NKOGMNSIRHNSLN 133  
 Db 46 nkqgwnsirhnsln 61

RESULT 11  
 R44551  
 ID R44551 standard; Protein; 480 AA.  
 XX  
 AC R44551;

DT 02-JUN-1994 (first entry)  
 XX  
 DE Brain factor-1.

XX Brain factor; BF-1; BF-2; BF-3; DNA binding domain;  
 KW brain transcription factor; diagnosis; tumour; cancer; probe;  
 telencephalon.  
 XX  
 OS Rattus rattus.

XX  
 FH Key Location/Qualifiers  
 FT Domain 162..271  
 FT /Label= DNA\_binding\_domain\_homology  
 FN W09323430-A.

XX  
 PD 25-NOV-1993.

XX  
 PE 30-APR-1993; 93MO-US04102.

XX  
 PR 13-MAY-1992; 92US-0882292.

XX  
 PA (SLOK ) SLOAN KETTERING INST CANCER.

XX  
 PL Lai E, Tao W;

XX  
 DR WPI: 1993-386481/48.  
 DR N-PSDB; Q50631.

XX  
 PT Isolated, animal nucleic acids encoding brain transcription  
 PT factors - useful for diagnosis and treatment of abnormal brain  
 PT factor synthesis in tumour tissue from animals and utilised as  
 PT probe

PS Claim 23; Page 57-60; 96pp; English.

CC The brain factors are transcription factors. Abnormal expression of  
 CC BF-1 in telencephalon-derived tissue or tumour tissue can be  
 CC diagnosed. BF DNA or protein can be used to correct defective  
 CC synthesis of BF.  
 XX  
 SO Sequence 480 AA;

Query Match 2.9%; Score 16; DB 14; Length 480;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 NKOGMNSIRHNSLN 133

Db 212 nkqgwnsirhnsln 227

RESULT 12  
 ID Y55737 standard; Protein; 76 AA.  
 XX  
 AC Y55737;

DT 22-FEB-2000 (first entry)  
 XX  
 DE FKHL10 forkhead domain partial fragment.

XX Forkhead transcription factor gene; FKHL7; treatment; glaucoma; human;  
 KW transgenic animal; drug screening.  
 XX  
 OS Unidentified.

XX  
 PN W09953060-A2.

PD 21-OCT-1999.

PE 14-APR-1999; 99MO-US08148.

PR 15-APR-1998; 98US-0081870.

PR 22-MAY-1998; 98US-0083352.

PA (IOWA ) UNIV IOWA RES FOUND.

PI Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;

DR WPI: 1999-620429/53.

XX  
 PT New isolated human forkhead transcription factor gene, FKHL7, used to,  
 PT e.g. develop products for the diagnosis -  
 PT  
 XX

PS Disclosure; Fig 2; 99pp; English.

CC The invention provides a human forkhead transcription factor gene, FKHL7.  
 CC The FKHL7 protein can be produced by standard recombinant methodology.  
 CC The products can be used for diagnosis, prognosis, monitoring, prevention  
 CC and treatment of glaucoma. They can also be used for the production of  
 CC transgenic animals and drug screening. Sequences Y55722-739 represent  
 CC forkhead domains of different members of the FKHL-family of genes.  
 XX

SO Sequence 76 AA;

Query Match 2.4%; Score 13; DB 20; Length 76;  
 Best Local Similarity 100.0%; Pred. No. 0.00057;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GWMNSIRHNSLN 133  
 Db 49 gwmsirhnsln 61

RESULT 13  
 Y55729  
 ID Y55729 standard; Protein; 106 AA.  
 XX  
 AC Y55729;

DT 22-FEB-2000 (first entry)  
 XX  
 DE FKHL9 forkhead domain fragment.

XX Forkhead transcription factor gene; FKHL7; treatment; glaucoma; human;  
 KW transgenic animal; drug screening.  
 XX  
 OS Unidentified.

PN W09953060-A2.  
 XX  
 PD 21-OCT-1999.  
 XX  
 PF 14-APR-1999; 99WO-US08148.  
 XX  
 PR 15-APR-1998; 98US-0081870.  
 XX 22-MAY-1998; 98US-0083352.  
 PA (IOWA ) UNIV IOWA RES FOUND.  
 XX  
 PI Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;  
 DR WPI: 1999-620429/53.  
 XX  
 PT New isolated human forkhead transcription factor gene, FKHL7, used to,  
 PT e.g. develop products for the diagnosis -  
 XX  
 PS Disclosure: Fig 2; 99pp; English.  
 CC The invention provides a human forkhead transcription factor gene, FKHL7.  
 CC The FKHL7 protein can be produced by standard recombinant methodology.  
 CC The products can be used for diagnosis, prognosis, monitoring, prevention  
 CC and treatment of glaucoma. They can also be used for the production of  
 CC transgenic animals and drug screening. Sequences Y5722-739 represent  
 CC forkhead domains of different members of the FKHL-family of genes.  
 CC  
 XX  
 SQ Sequence 106 AA;

Query Match 2.4%; Score 13; DB 20; Length 106;  
 Best Local Similarity 100.0%; Pred. No. 0.00076;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 121 GWONSIRHNSLN 133  
 |||||  
 Db 49 gwqnsirhnsln 61

RESULT 14  
 W03762  
 ID W03762 standard; Protein: 540 AA.  
 XX  
 AC W03762;  
 XX  
 DT 31-OCT-1996 (first entry)  
 XX  
 DE Interleukin binding factor ILF-1.  
 XX  
 KW Interleukin binding factor; ILF-1; purine rich motif; HIV-LTR;  
 KW human immunodeficiency virus; long terminal repeat; fork head;  
 KW AIDS; DNA binding; nuclear factor of activated T cell; NFAT;  
 KW autoimmune disease; rheumatism.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Binding-site 96..101  
 FT /label= nucleotide-binding\_site  
 FT /note= "putative; is homologous to sequences  
 FT found in a number of enzymes including  
 FT the ras oncogene"  
 FT  
 FT Region 228..233  
 FT /label= nuclear\_localisation\_signal  
 FT /note= "putative"  
 FT  
 FT Region 239..247  
 FT /label= ubiquitin-mediated\_degradation\_signal  
 FT /note= "putative"  
 FT  
 FT Modified-site 458..460  
 FT /label= N-glycosylation\_site  
 FT /note= "putative"  
 FT 136..233  
 FT /note= "homologous to Drosophila fork head DNA-  
 FT Domain

binding domain"  
 XX  
 PN US5534631-A.  
 XX  
 PD 09-JUL-1996.  
 XX  
 PF 30-JUN-1992; 92US-0906930.  
 XX  
 PR 30-JUN-1992; 92US-0906930.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 XX  
 PI Gaynor RB, Li C, Nirula A;  
 DR WPI: 1996-333310/33.  
 DR N-PSDB: T36079.  
 XX  
 PT New DNA encoding interleukin (IL) nucleic acid binding protein -  
 PT that binds to purine-rich motifs in viral and IL gene promoters to  
 PT modulate their expression, e.g. for treating AIDS and auto-immune  
 PT disease  
 PS  
 PS Claim 3: Columns 29-32; 61pp; English.  
 XX  
 CC The present sequence is that of interleukin binding factor ILF-1,  
 CC which binds specifically to a purine-rich motif (the NFAT binding  
 CC site) in the long terminal repeat of HIV and, with lower affinity,  
 CC to a similar sequence in the interleukin (IL)-2 promoter. Database  
 CC searches have shown that a 98 amino acid region of ILF-1 is  
 CC homologous to the Drosophila fork head regulatory protein, and the  
 CC DNA binding domain of the hepatocyte specific factor HNF-3A. The  
 CC ILF regulates IL-2 expression in T lymphocytes and is  
 CC involved in positive and negative regulation of viral and cellular  
 CC genes; it is thus potentially useful for treating AIDS and to  
 CC inhibit lymphocytes in rheumatic and other autoimmune diseases.  
 CC A second ILF protein, designated ILF-2, has also been identified.  
 CC The two proteins are identical up to amino acid 475 and then differ  
 CC in their C-terminal sequences.  
 CC  
 XX  
 SQ Sequence 540 AA;

Query Match 2.4%; Score 13; DB 17; Length 540;  
 Best Local Similarity 100.0%; Pred. No. 0.0032;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 121 GWONSIRHNSLN 133  
 |||||  
 Db 181 gwqnsirhnsln 193

RESULT 15  
 Y5730  
 ID Y5730 standard; Protein: 106 AA.  
 XX  
 AC Y5730;  
 XX  
 DT 22-FEB-2000 (first entry)  
 XX  
 DE FKHL8 forkhead domain fragment.  
 XX  
 KW Forkhead transcription factor gene; FKHL7; treatment; glaucoma; human;  
 KW transgenic animal; drug screening.  
 XX  
 OS unidentified.  
 XX  
 PN W09953060-A2.  
 XX  
 PD 21-OCT-1999.  
 XX  
 PF 14-APR-1999; 99WO-US08148.  
 XX  
 PR 15-APR-1998; 98US-0081870.

PR 22-MAY-1998; 98US-0083352.  
XX  
XX  
PA (IOWA ) UNIV IOWA RES FOUNO.  
XX  
XX  
PI Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;  
XX  
XX WPI. 1999-620429/53.  
DR  
XX  
XX New isolated human forkhead transcription factor gene, FKHL7, used to,  
PT e.g. develop products for the diagnosis -  
XX  
XX  
PS Disclosure: Fig 2; 99pp; English.  
XX  
XX  
CC The invention provides a human forkhead transcription factor gene, FKHL7.  
CC The FKHL7 protein can be produced by standard recombinant methodology.  
CC The products can be used for diagnosis, prognosis, monitoring, prevention  
CC and treatment of glaucoma. They can also be used for the production of  
CC transgenic animals and drug screening. Sequences Y55722-739 represent  
CC forkhead domains of different members of the FKHL-family of genes.  
XX  
XX  
SQ Sequence 106 AA;

Query Match 2.2%; Score 12; DB 20; Length 106;  
Best Local Similarity 100.0%; Pred. No. 0.0056;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 77 VKPPYSTIALIT 88  
|||  
Db 5 vkppysyalit 16

Search completed: February 6, 2001, 22:52:58  
Job time: 8827 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 6, 2001, 21:50:24 ; Search time 60.6 Seconds  
(without alignments)  
1069.571 Million cell updates/sec

Title: US-09-292-862-2

Perfect score: 553  
Sequence: 1 MQARYSVSPNSLGVPYILG.....PSSQSLVYRTSGAFVYDCSKF 553

Scoring table: OLIGO  
Gapop 60.0, Gapext 60.0

Searched: 374700 seqs, 117207915 residues

Word size: 0

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

SPREMBL\_15:  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.protist:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	351	63.5	553	4 Q9NUE5	Q9NUE5 homo sapien
2	120	21.7	553	11 Q9QWR9	Q9QWR9 mus musculu
3	69	12.5	528	13 Q93440	Q93440 gallus gall
4	68	12.3	492	13 Q9YHB2	Q9YHB2 xenopus lae
5	68	12.3	492	13 Q9PVZ3	Q9PVZ3 xenopus lae
6	61	11.0	445	13 Q9Y771	Q9Y771 gallus gall
7	61	11.0	445	13 Q13017	Q13017 gallus gall
8	40	7.2	461	13 Q9PVY8	Q9PVY8 xenopus lae
9	40	7.2	465	13 Q9PVY9	Q9PVY9 xenopus lae
10	39	7.1	465	13 Q9DDO0	Q9DDO0 homo sapien
11	21	3.8	108	4 Q22510	Q22510 caenorhabd
12	19	3.4	270	5 Q9VOV5	Q9VOV5 drosophila
13	16	2.9	62	13 Q9PS85	Q9PS85 xenopus xf
14	16	2.9	90	11 Q88470	Q88470 mus musculu
15	16	2.9	365	5 Q9VZK0	Q9VZK0 drosophila
16	16	2.9	402	5 Q61733	Q61733 brachydanio
17	16	2.9	420	13 Q73862	Q73862 brachydanio
18	16	2.9	436	13 Q9YHCS	Q9YHCS xenopus lae
19	15	2.7	202	5 P91278	P91278 caenorhabd

20	15	2.7	321	13 Q73784	Q73784 brachydanio
21	15	2.7	323	5 Q10924	Q10924 caenorhabd
22	15	2.7	352	13 Q9PRJ8	Q9PRJ8 xenopus lae
23	15	2.7	353	13 Q73785	Q73785 brachydanio
24	15	2.7	353	13 Q9PR68	Q9PR68 xenopus lae
25	15	2.7	371	13 Q73782	Q73782 brachydanio
26	15	2.7	443	13 P79770	P79770 gallus gall
27	15	2.7	456	5 Q9W1Y5	Q9W1Y5 drosophila
28	15	2.7	492	11 Q35392	Q35392 mus musculu
29	15	2.7	492	11 Q9NU39	Q9NU39 homo sapien
30	14	2.5	408	4 P91587	P91587 caenorhabd
31	13	2.4	284	4 P91587	P91587 drosophila
32	13	2.4	322	5 Q9VOV4	Q9VOV4 xenopus lae
33	13	2.4	345	13 Q9PSY4	Q9PSY4 xenopus lae
34	13	2.4	367	13 Q91905	Q91905 xenopus lae
35	13	2.4	370	13 Q91904	Q91904 xenopus lae
36	13	2.4	409	11 Q35939	Q35939 mus musculu
37	13	2.4	442	5 Q91133	Q91133 brachydanio
38	13	2.4	443	4 Q9UJL0	Q9UJL0 homo sapien
39	13	2.4	447	13 Q73628	Q73628 anolis caro
40	13	2.4	517	3 Q13606	Q13606 schizosach
41	13	2.4	586	3 P79006	P79006 schizosach
42	13	2.4	605	5 Q9YTF7	Q9YTF7 drosophila
43	13	2.4	609	4 Q13623	Q13623 homo sapien
44	13	2.4	642	3 Q60129	Q60129 schizosach
45	13	2.4	663	3 Q43058	Q43058 schizosach

## ALIGNMENTS

RESULT 1  
ID Q9NUE5 PRELIMINARY: PRT: 553 AA.  
AC Q9NUE5  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE D118B18.1 (FORKHEAD BOX C1).  
GN FOXO1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Patel R.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AI034344; CAB81658.1;  
DR SQUENCE 553 AA; 56788 MW; 59C6FB94303ED59A CRC64;  
SQ

Query Match	63.5%;	Score 351;	DB 4;	Length 553;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 351;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	203	PAPPEQADGNAPGPPPPVRIODIKTEGTCSPQPLSPAAALGSGSAAAVPKIESPDS	262	
DB	203	PAPPEQADGNAPGPPPPVRIODIKTEGTCSPQPLSPAAALGSGSAAAVPKIESPDS	262	
QY	263	SSSSLSGSSSPGSLSPARPLSLDGADSDAPPPAPAPPPHHSOGFSDVNTSLRGSPQ	322	
DB	263	SSSSLSGSSSPGSLSPARPLSLDGADSDAPPPAPAPPPHHSOGFSDVNTSLRGSPQ	322	
QY	323	SAAEELSSGLLAASAAASRAGIAPPLALGAYSPGSSIVSPSSQTSAGSSGGGGGAG	382	
DB	323	SAAEELSSGLLAASAAASRAGIAPPLALGAYSPGSSIVSPSSQTSAGSSGGGGGAG	382	
QY	383	AAGCAGGAGTYHCNLOAMSLYAGERGSHLQAGAGGASAVDDPLPDVSLPPVSSSSS	442	
DB	383	AAGCAGGAGTYHCNLOAMSLYAGERGSHLQAGAGGASAVDDPLPDVSLPPVSSSSS	442	
QY	443	SLSHGGGGGGGGGGGAGHHPAAGGRLTSWYLNQAGDLGHLASAAAAAAGYPGQOQ	502	

DB 443 SLHGGGGGGGGGEGAGHHPAAGRLTSLWYNAGGDLHLASAAAAAAGYPCOOQ 502  
OY 503 NHHVREMEFESORIGLNNSPVNGNSCOMAFSSOSLYRTSGAFVYDCSKF 553  
DB 503 NHHVREMEFESORIGLNNSPVNGNSCOMAFSSOSLYRTSGAFVYDCSKF 553

## RESULT 2

O9QWR9

ID O9QWR9 PRELIMINARY; PRT; 553 AA.

AC O9QWR9; PRT; 553 AA.

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DE 01-OCT-2000 (TREMBlrel. 15, Last annotation update)

GN FKHL/ME1

OS Mus musculus (Mouse)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=EMBRYO;

RA Hiemisch H., Schütz G., Kaestner K.H.;

RT "The mouse ~~phosphatase~~ gene: cDNA sequence, chromosomal localization and

expression in adult tissues."

RL EMBL: AJ23298; CA11239.1; "

HSSP: Q63245; 2HFH.

DR INTERPRO: IPR001766;

DR PFAM: PF00250; Fork head; 1.

DR PRINTS: PR00053; FORKHEAD.

DR PROSITE: PS00657; FORK\_HEAD\_1; 1.

DR PROSITE: PS00658; FORK\_HEAD\_2; 1.

DR PROSITE: PS50039; FORK\_HEAD\_3; 1.

SQ SEQUENCE 553 AA; 56939 MW; 3CDB14F69AA2F217 CRC64;

## Query Match

Best Local Similarity 100.0%; Score 120; DB 11; Length 553;

Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 MARAGPTPOPQPKDMWKKPYSYIALITMAIQAPPKKITLNGYOFIMDRPFPRDK 119

DB 60 MARAGPTPOPQPKDMWKKPYSYIALITMAIQAPPKKITLNGYOFIMDRPFPRDK 119

OY 120 OGWONSIRHNLSTNECFVKKVPRDDKKPGKGSYWTLPDPSYNNFENGSLRRRFRKKDA 179

DB 120 OGWONSIRHNLSTNECFVKKVPRDDKKPGKGSYWTLPDPSYNNFENGSLRRRFRKKDA 179

## RESULT 3

O93440

ID O93440 PRELIMINARY; PRT; 528 AA.

AC O93440; PRT; 528 AA.

DT 01-NOV-1998 (TREMBlrel. 08, Created)

DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)

DE 01-OCT-2000 (TREMBlrel. 15, Last annotation update)

GN WINGED-HELIX TRANSCRIPTION FACTOR.

OS Gallus gallus (Chicken)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OX NCBI\_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=TOTAL EMBRYO;

RX MEDLINE=98264590; PubMed=9603427.

RA Buchberger A., Schwarzer M., Brandt T., Pabst O., Seidl K.,

RT "Chicken winged-helix transcription factor CFKH-1 prefigures axial and

RT appendicular skeletal structures during chicken embryogenesis."

RL Dev. Dyn. 212:94-101(1998).

DR EMBL: Y17794; CA676851.1; "

DR HSSP: Q63245; 2HFH.

DR INTERPRO: IPR001766; "

DR PFAM: PF00250; Fork head; 1.

DR PRINTS: PR00053; FORKHEAD.

DR PROSITE: PS00657; FORK\_HEAD\_1; 1.

DR PROSITE: PS00658; FORK\_HEAD\_2; 1.

DR PROSITE: PS50039; FORK\_HEAD\_3; 1.

SQ SEQUENCE 528 AA; 54830 MW; 87FD601D05F70128 CRC64;

Query Match

Best Local Similarity 100.0%; Score 69; DB 13; Length 528;

Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 RPFYRDNRKQWONSIRHNLSTNECFVKKVPRDDKKPGKGSYWTLPDPSYNNFENGSLRR 170

DB 72 RPFYRDNRKQWONSIRHNLSTNECFVKKVPRDDKKPGKGSYWTLPDPSYNNFENGSLRR 131

OY 171 RRRFRKKDA 179

DB 132 RRRFRKKDA 140

## RESULT 4

O9YHB2

ID O9YHB2 PRELIMINARY; PRT; 492 AA.

AC O9YHB2; PRT; 492 AA.

DT 01-MAY-1999 (TREMBlrel. 10, Created)

DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)

DE WINGED HELIX TRANSCRIPTION FACTOR XFD-11.

OS Xenopus laevis (African clawed frog)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

OX Xenopodinae; Xenopus.

RN NCBI\_TaxID=8335;

RP SEQUENCE FROM N.A.

RC MEDLINE=98440281; PubMed=9767159;

RA Koster M., Dillinger K., Knoche W.;

RT "Expression pattern of the winged helix factor XFD-11 during Xenopus

embryogenesis."

RL Mech. Dev. 76:169-173(1998).

DR EMBL: AF116844; AAC9469.1; "

DR HSSP: Q63245; 2HFH.

DR INTERPRO: IPR001766; "

DR PFAM: PF00250; Fork head; 1.

DR PRINTS: PR00053; FORKHEAD.

DR PROSITE: PS00657; FORK\_HEAD\_1; 1.

DR PROSITE: PS00658; FORK\_HEAD\_2; 1.

DR PROSITE: PS50039; FORK\_HEAD\_3; 1.

SQ SEQUENCE 492 AA; 53643 MW; A6C66458CF44465F CRC64;

Query Match

Best Local Similarity 100.0%; Score 68; DB 13; Length 492;

Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 RPFYRDNRKQWONSIRHNLSTNECFVKKVPRDDKKPGKGSYWTLPDPSYNNFENGSLRR 170

DB 112 RPFYRDNRKQWONSIRHNLSTNECFVKKVPRDDKKPGKGSYWTLPDPSYNNFENGSLRR 171

OY 171 RRRFRKKD 178

DB 172 RRRFRKKD 179

## RESULT 5

O9PVZ3

ID O9PVZ3 PRELIMINARY; PRT; 492 AA.

AC 09PV23; 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE XFD-11 PROTEIN.  
 GN XFD-11.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OC NCBI\_TaxID=8335;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP Knoechel W.;  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA 12)  
 RT "Expression pattern of the winged helix factor XFD-11 during Xenopus  
 embryogenesis.";  
 RL Mech. Dev. 76:169-173(1998).  
 DR EMBL; AJ242675; CAB44727.1;  
 DR HSSP; Q63245; 2HFH.  
 DR INTERPRO; IPR001766;  
 DR PFM; PF00250; Fork\_head; 1.  
 DR PRINTS; PR00053; FORKHEAD.  
 DR PROSITE; PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE; PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE; PS50039; FORK\_HEAD\_3; 1.  
 DR PROSITE; PS50039; FORK\_HEAD\_3; 1.  
 SQ SEQUENCE 492 AA; 53674 MW; 66906A5A9A2F4E CRC64;

Query Match 12.3%; Score 68; DB 13; Length 492;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-52;  
 Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 RFFYDNNQGNQNSIRHNLSTNECFVAVPRDDKPKGSGSYWTLDPDSYNNFENGSLRR 170  
 112 RFFYDNNQGNQNSIRHNLSTNECFVAVPRDDKPKGSGSYWTLDPDSYNNFENGSLRR 171  
 DB 112 RFFYDNNQGNQNSIRHNLSTNECFVAVPRDDKPKGSGSYWTLDPDSYNNFENGSLRR 171  
 OY 171 RRRFKKD 178  
 172 RRRFKKD 179  
 DB 172 RRRFKKD 179

RESULT 6  
 P79771 PRELIMINARY; PRT; 445 AA.  
 AC 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE WINGED HELIX PROTEIN CMH-2.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OC NCBI\_TaxID=9031;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE-EMBRYO.  
 RC MEDLINE-97141794; PubMed-6988052;  
 RA Freyaldenhoven B.S., Freyaldenhoven M.P., Iacovoni J.S., Vogt P.K.;  
 RT "Aberrant cell growth induced by avian winged helix proteins.";  
 RL Cancer Res. 57:123-129(1997).  
 DE EMBL; U37273; AAC60065.1;  
 DR HSSP; Q63245; 2HFH.  
 DR INTERPRO; IPR001766;  
 DR PFM; PF00250; Fork\_head; 1.  
 DR PRINTS; PR00053; FORKHEAD.  
 DR PROSITE; PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE; PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE; PS50039; FORK\_HEAD\_3; 1.

SQ SEQUENCE 445 AA; 48067 MW; 5F32D5CE521113F CRC64;

Query Match 11.0%; Score 61; DB 13; Length 445;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-46;  
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 118 NKQGMQNSIRHNLSTNECFVAVPRDDKPKGSGSYWTLDPDSYNNFENGSLRRRRRFRKK 177  
 79 NKQGMQNSIRHNLSTNECFVAVPRDDKPKGSGSYWTLDPDSYNNFENGSLRRRRRFRKK 138  
 DB 178 D 178  
 DB 139 D 139

RESULT 7  
 O13017 PRELIMINARY; PRT; 468 AA.  
 AC 01-JUL-1997 (TREMBLrel. 04, Created)  
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE WINGED HELIX TRANSCRIPTIONAL FACTOR MRF-1.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OC NCBI\_TaxID=9031;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP Arase Y., Koseki H., Miura N.;  
 RA Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; U95823; AAB53749.1;  
 DR EMBL; U95823; AAB53749.1;  
 DR HSSP; Q63245; 2HFH.  
 DR INTERPRO; IPR001766;  
 DR PFM; PF00250; Fork\_head; 1.  
 DR PRINTS; PR00053; FORKHEAD.  
 DR PROSITE; PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE; PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE; PS50039; FORK\_HEAD\_3; 1.  
 DR PROSITE; PS50039; FORK\_HEAD\_3; 1.  
 SQ SEQUENCE 468 AA; 50667 MW; EAFCA57ADF311746 CRC64;

Query Match 11.0%; Score 61; DB 13; Length 468;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-46;  
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 118 NKQGMQNSIRHNLSTNECFVAVPRDDKPKGSGSYWTLDPDSYNNFENGSLRRRRRFRKK 177  
 112 NKQGMQNSIRHNLSTNECFVAVPRDDKPKGSGSYWTLDPDSYNNFENGSLRRRRRFRKK 171  
 DB 178 D 178  
 DB 172 D 172

RESULT 8  
 O9PVY8 PRELIMINARY; PRT; 461 AA.  
 AC 09PVY8; 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE FD-4' PROTEIN.  
 GN Xenopus laevis (African clawed frog).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OC NCBI\_TaxID=8335;  
 OX [1]  
 RN SEQUENCE FROM N.A.

RA Kostler M., Dillinger K., Knochel W.;  
RT "Xenopus winged helix factors XFD-4/4" represent orthologues to  
RT mammalian MFH-1";  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AJ249225; CAB54144.1; -  
DR HSSP: 063245; 2HFH.  
DR INTERPRO: IPR001766; -  
DR PFAM: PF00250; Fork\_head; 1.  
DR PRINTS: PR00053; FORKHEAD.  
DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
DR PROSITE: PS00659; FORK\_HEAD\_3; 1.  
SQ SEQUENCE 461 AA; 51162 MW; DE672C9F715BCF71 CRC64;

Query Match  
Best Local Similarity 7.2%; Score 40; DB 13; Length 461;  
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 VKPPSYIALITMAIONAPDKKITTNGIYOFIMDRPFYR 116  
Db 71 VKPPSYIALITMAIONAPDKKITTNGIYOFIMDRPFYR 110

RESULT 9  
Q9PVY9 PRELIMINARY; PRT; 465 AA.  
AC Q9PVY9;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
GN FD-4 PROTEIN.

OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
OX NCBI\_TaxID=8353;  
RN [1]  
RP SEQUENCE FROM N.A.

RA Kostler M., Dillinger K., Knochel W.;  
RT "Xenopus winged helix factors XFD-4/4" represent orthologues to  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ249224; CAB54143.1; -  
DR HSSP: 063245; 2HFH.  
DR INTERPRO: IPR001766; -  
DR PFAM: PF00250; Fork\_head; 1.  
DR PRINTS: PR00053; FORKHEAD.  
DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
DR PROSITE: PS00659; FORK\_HEAD\_3; 1.  
SQ SEQUENCE 465 AA; 51613 MW; 15DFE9AD881E7245 CRC64;

Query Match  
Best Local Similarity 7.2%; Score 40; DB 13; Length 465;  
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 VKPPSYIALITMAIONAPDKKITTNGIYOFIMDRPFYR 116  
Db 71 VKPPSYIALITMAIONAPDKKITTNGIYOFIMDRPFYR 110

RESULT 10  
Q9UDDO PRELIMINARY; PRT; 108 AA.  
AC Q9UDDO;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
FOKHEAD TRANSCRIPTIONAL ACTIVATOR HOMOLOG (FRAGMENT).  
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OX Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE-93271467; PubMed=849623;  
RA Thomas R., Moore J., Johnston T., Socha C., Klemz M.;  
RT "Drosophila forkhead homologues are expressed in a lineage-restricted  
manner in human hematopoietic cells";  
RL Blood 81:2854-2859(1993).  
DR HSSP: 063245; 2HFH.  
DR INTERPRO: IPR001766; -  
DR PFAM: PF00250; Fork\_head; 1.  
DR PRINTS: PR00053; FORKHEAD.  
DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
DR PROSITE: PS00659; FORK\_HEAD\_3; 1.  
SQ SEQUENCE 108 AA; 12886 MW; A6A1511F7D2899B2 CRC64;

Query Match  
Best Local Similarity 7.1%; Score 39; DB 4; Length 108;  
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 PFYRDNKQGMQNSIRHNLSTNECFVAVPRDCKPKGKSY 151  
Db 44 PFYRDNKQGMQNSIRHNLSTNECFVAVPRDCKPKGKSY 82

RESULT 11  
Q22510 PRELIMINARY; PRT; 270 AA.  
AC Q22510;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
GN T14G12.4.  
OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OX Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE-94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Jones M., Kershaw J., Kirsten J., Lalster N., Jier M., Johnston L.,  
RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rinken L., Roopra A., Saunders D., Showkeen R.,  
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Wellstock L., Wilkinson-Sproat J., Woldman P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans";  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.

RA Wilcox L.;  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
RP SEQUENCE FROM N.A.

RA Waterston R.;  
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U41268; AA82436.1; -  
DR HSSP: 063245; 2HFH.  
DR INTERPRO: IPR001766; -  
DR PFAM: PF00250; Fork\_head; 1.  
DR PRINTS: PR00053; FORKHEAD.  
DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
DR PROSITE: PS00658; FORK\_HEAD\_2; 1.

DR PROSITE: PS50039; FORK\_HEAD\_3; 1.  
SQ SEQUENCE 270 AA; 30491 MW; 7C49116E5EC76175 CRC64;

Query Match 3.8%; Score 21; DB 5; Length 270;  
Best Local Similarity 100.0%; Pred. No. 1,1e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 113 YRDNGKQGMNSIRHNLSTLN 133  
Db 128 PFYRKNKQGMNSIRHNLSTLN 148

RESULT 12  
O9VOV5 PRELIMINARY: PRT: 451 AA.  
AC O9VOV5;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)  
DE SLP2. PROTEIN.  
GN SLP2.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_Taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=BERKELEY;  
MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
Abri J.F., Agbayani A., An H.-J., Andrews-Plamkoc C., Baldwin D.,  
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
Borkova D., Botchan M.R., Bouck J., Brostein P., Brotler P.,  
Burtis K.C., Busam D.A., Butler H., Davenport L.B., Davies P.,  
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies S.M.,  
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunov B.C., Dunn P.,  
Durtin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
Foster C., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
Merklov G., Milshina N.V., Moberly L., Muzny D.M., Nelson D.L.,  
Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
Nelson D.R., Nelson K.A., Nixon K., Pollard J.D., Puri V., Reese M.G.,  
Palazzo M., Pittman G.S., Pan S., Pollard J.D., Schaefer F., Shen H.,  
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith T.,  
Shue B.C., Siden-Klamos I., Simpson M., Skupski R., Sun E.,  
Slyter E., Spradling A.C., Turner R., Venter E., Wang A.H., Wang X.,  
Syrakas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
Wang Z.-T., Wasserman D.A., Weinstein G.M., Weissbach J.,  
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao S., Zhu X., Smith H.O.,  
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of Drosophila melanogaster."  
RL Science 287:2185-2195(2000).  
DR EMBL: AE003578; AAF51057.1;  
DR HSSP: Q63245; 2HRH.  
DR FLYBASE: FBgn0004567; slp2.  
DR INTERPRO: IPR000087; -.

DR INTERPRO: IPR001766; -  
DR PFAM: PF00250; FORK\_head; 1.  
DR PRINTS: PR00053; FORKHEAD.  
DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
DR PROSITE: PS50039; FORK\_HEAD\_3; 1.  
SQ SEQUENCE 451 AA; 49237 MW; 170C7338503B86B5 CRC64;

Query Match 3.4%; Score 19; DB 5; Length 451;  
Best Local Similarity 100.0%; Pred. No. 1e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 115 YRDNGKQGMNSIRHNLSTLN 133  
Db 217 YRDNGKQGMNSIRHNLSTLN 235

RESULT 13  
O9PS85 PRELIMINARY: PRT: 62 AA.  
ID O9PS85;  
AC O9PS85;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)  
DE XFKH4 FORK HEAD PROTEIN (FRAGMENT).  
GN Xenopus.  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
OC Xenopodinae.  
OX NCBI\_Taxid=8353;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96083056; PubMed=7586752;  
RA Diksen M.L., Jamrich M.;  
RT "Differential expression of fork head genes during early Xenopus and  
zebrafish development."  
RL dev. Genet. 17:107-116(1995).  
DR HSSP: Q63245; 2HRH.  
DR INTERPRO: IPR001766; -  
DR PFAM: PF00250; FORK\_head; 1.  
DR PRINTS: PR00053; FORKHEAD.  
DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
DR PROSITE: PS50039; FORK\_HEAD\_3; 1.  
SQ SEQUENCE 62 AA; 7468 MW; D2DA711E65AE70AF CRC64;

Query Match 2.9%; Score 16; DB 13; Length 62;  
Best Local Similarity 100.0%; Pred. No. 9.9e-07;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 118 NKQGMNSIRHNLSTLN 133  
Db 41 NKQGMNSIRHNLSTLN 56

RESULT 14  
O88470 PRELIMINARY: PRT: 90 AA.  
ID O88470;  
AC O88470;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)  
DE PITUITARY FORKHEAD FACTOR (FRAGMENT).  
GN PPRK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98283916; PubMed=9620855;

RA Treier M., Gleiberman A.S., O'Connell S.M., Szeto D.P., McMahon J.A.,  
 RA McMahon A.P., Rosenfeld M.G.,  
 RT "Multistep signaling requirements for pituitary organogenesis in  
 RT vivo.",  
 RL Genes Dev. 12:1691-1704(1998).  
 DR EMBL: AF060873; AAC27508.1; -.  
 DR HSSP: O63245; 2HFH.  
 DR MGD: MGI:1349428; PERK.  
 DR INTERPRO: IPR001766; -.  
 DR PFAM: PF00250; Fork\_head; 1.  
 DR PRINTS: PR00053; FORKHEAD.  
 DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE: PS50039; FORK\_HEAD\_3; 1.  
 FT NON\_TER 1 1  
 FT TER 90 90  
 SQ SEQUENCE 90 AA; 10484 MW; 03CEDDEC6BA84BEB CRC64;

Query Match 2.9%; Score 16; DB 11; Length 90;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GMONSTRHNLNECF 136  
 Db 44 GMONSTRHNLNECF 59

RESULT 15  
 Q9VZK0 PRELIMINARY; PRT; 365 AA.  
 AC Q9VZK0;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE FD64A PROTEIN.  
 GN FD64A.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_taxid=7227;  
 RX 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abriel J.F., Agbayani A., An H.-J., Andrews-Plankoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bertman B.P., Brokstein P., Brotlier P.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brokstein P.,  
 RA Burlis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Degen Z., Dey A.D., Dew I., Dietz S.M.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostli D., Houston K.A., Howland T.J., Mei M.-H., Ibegyan C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svilaras R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Williams S.M., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao O., Zheng L.,  
 RA Zhong X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003480; AAF47821.1; -.  
 DR HSSP: O63245; 2HFH.  
 DR FLYBASE: FBgn0004895; fd64A.  
 DR INTERPRO: IPR001766; -.  
 DR PFAM: PF00250; Fork\_head; 1.  
 DR PRINTS: PR00053; FORKHEAD.  
 DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE: PS50039; FORK\_HEAD\_3; 1.  
 SQ SEQUENCE 365 AA; 40635 MW; 4FCE959A285FCA0F CRC64;

Query Match 2.9%; Score 16; DB 5; Length 365;  
 Best Local Similarity 100.0%; Pred. No. 4.1e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

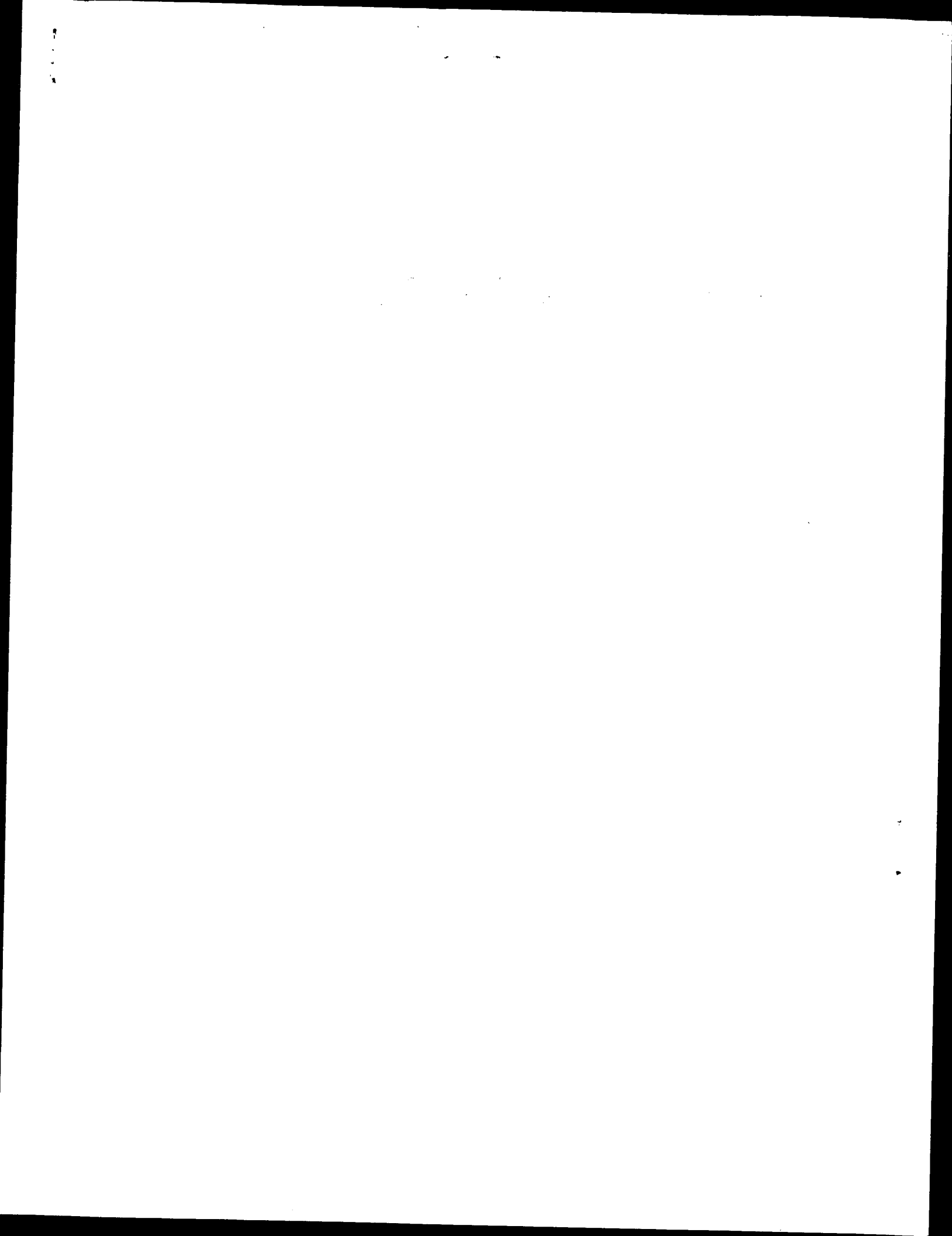
QY 118 NKGWONSTRHNLN 133  
 Db 131 NKGWONSTRHNLN 146

Search completed: February 6, 2001, 23:38:53  
 Job time: 6509 sec

Wed Feb 7 14:06:00 2001

us-09-292-862-2.rspt

---





Wed Feb 7 14:05:57 2001

us-09-292-862-2.ra1

Page 1

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 6, 2001, 20:37:49 ; Search time 43.35 Seconds  
(Without alignments)  
229.071 Million cell updates/sec

Title: US-09-292-862-2  
Perfect score: 553  
Sequence: 1 MQARVSVSPNSLGVVPIYL.....PSQSLXRTSGAFVYDCSKF 553

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 174772 seqs, 17957048 residues

Word size : 0 174772

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :  
1: /cgn2\_6/prodata/2/iaa/5A.COMB.pep:\*  
2: /cgn2\_6/prodata/2/iaa/5B.COMB.pep:\*  
3: /cgn2\_6/prodata/2/iaa/5.COMB.pep:\*  
4: /cgn2\_6/prodata/2/iaa/PCITUS.COMB.pep:\*  
5: /cgn2\_6/prodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	250	45.2	553	3 US-09-083-351-2	Sequence 2, Appl1
2	106	19.2	106	3 US-09-083-351-5	Sequence 5, Appl1
3	61	11.0	106	3 US-09-083-351-6	Sequence 6, Appl1
4	23	4.2	106	3 US-09-083-351-7	Sequence 7, Appl1
5	16	2.9	106	3 US-09-083-351-8	Sequence 8, Appl1
6	16	2.9	106	3 US-09-083-351-16	Sequence 16, Appl1
7	16	2.9	106	3 US-09-083-351-17	Sequence 17, Appl1
8	16	2.9	480	1 US-07-882-292-2	Sequence 2, Appl1
9	16	2.9	480	2 US-08-331-644-2	Sequence 2, Appl1
10	16	2.9	480	4 PCT-US93-04102-2	Sequence 12, Appl1
11	15	2.7	106	3 US-09-083-351-12	Sequence 13, Appl1
12	15	2.7	106	3 US-09-083-351-13	Sequence 13, Appl1
13	13	2.4	76	3 US-09-083-351-19	Sequence 19, Appl1
14	13	2.4	98	1 US-07-906-930E-4	Sequence 4, Appl1
15	13	2.4	106	3 US-09-083-351-11	Sequence 11, Appl1
16	13	2.4	540	1 US-07-906-930E-2	Sequence 2, Appl1
17	12	2.2	106	3 US-09-083-351-20	Sequence 20, Appl1
18	11	2.0	37	1 US-08-425-069-35	Sequence 35, Appl1
19	11	2.0	37	1 US-08-425-069-36	Sequence 36, Appl1
20	11	2.0	37	1 US-08-209-747-46	Sequence 46, Appl1
21	11	2.0	37	1 US-08-209-747-47	Sequence 47, Appl1
22	11	2.0	37	1 US-08-458-298-46	Sequence 46, Appl1
23	11	2.0	37	1 US-08-458-298-47	Sequence 47, Appl1
24	11	2.0	37	2 US-08-317-844B-35	Sequence 35, Appl1
25	11	2.0	37	2 US-08-317-844B-36	Sequence 36, Appl1
26	11	2.0	595	2 US-08-425-069-4	Sequence 4, Appl1
27	11	2.0	595	2 US-08-317-844B-4	Sequence 4, Appl1
28	11	2.0	4472	2 US-08-804-227C-2	Sequence 2, Appl1

29	11	2.0	4545	2 US-08-804-227C-14	Sequence 14, Appl1
30	10	1.8	10	1 US-07-969-307A-20	Sequence 20, Appl1
31	10	1.8	18	2 US-08-528-129A-10	Sequence 10, Appl1
32	10	1.8	20	1 US-08-234-602-14	Sequence 14, Appl1
33	10	1.8	35	1 US-08-209-747-52	Sequence 52, Appl1
34	10	1.8	35	1 US-08-458-298-52	Sequence 52, Appl1
35	10	1.8	98	1 US-07-906-930E-6	Sequence 9, Appl1
36	10	1.8	106	3 US-09-083-351-9	Sequence 9, Appl1
37	10	1.8	106	3 US-09-083-351-10	Sequence 10, Appl1
38	10	1.8	106	3 US-09-083-351-14	Sequence 14, Appl1
39	10	1.8	106	3 US-09-083-351-15	Sequence 15, Appl1
40	10	1.8	148	1 US-08-207-904-15	Sequence 15, Appl1
41	10	1.8	161	2 US-08-581-528A-6	Sequence 6, Appl1
42	10	1.8	161	4 PCT-US94-07799-6	Sequence 6, Appl1
43	10	1.8	199	3 US-08-506-553C-8	Sequence 8, Appl1
44	10	1.8	208	2 US-08-935-886-8	Sequence 8, Appl1
45	10	1.8	211	2 US-08-935-886-10	Sequence 10, Appl1

#### ALIGNMENTS

RESULT 1  
US-09-083-351-2  
Sequence 2, Application US/09083351  
Patent No. 6087107  
GENERAL INFORMATION:  
APPLICANT: Sheffield, Val C.  
APPLICANT: Alward, Wallace L.M.  
APPLICANT: Stone, Edwin M.  
APPLICANT: Nishimura, Darryl  
APPLICANT: Patil, Shiva  
TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
TITLE OF INVENTION: TRANSCRIPTION FACTOR  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: FOLEY, HONG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,351  
FILING DATE: 22-MAY-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: UTA-029,02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEO ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 553 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-083-351-2

Query Match 45.2% Score 250; DB 3; Length 553;  
Best Local Similarity 99.7%; Pred. No. 2.7e-215;  
Matches 350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Oy 203 PAPEQADGNAPGPPVPRVRIQIDIKTENGTCSPFPPLSPAAALSGSGAAVPIKIESPDS 262

|||||  
Db 203 PAPPEADGNAGPQPPPRIDIKTENGICSPRPPLSPAALICSSAAAVPRTIESPDS 262  
OY 263 SSSSSSSGSPSPGLPSARPLSLDGDADAPPPAPADPPPHHSGFVNDIMTSLRSGPQ 322  
Db 263 SSSSSSSGSPSPGLPSARPLSLDGDADAPPPAPADPPPHHSGFVNDIMTSLRSGPQ 322  
OY 323 SAAAEISSGLLASAAASSRAGIAPPLAGVSPGSSSYSPGSSSYSPGSSSYSPGSSSY 382  
Db 323 SAAAEISSGLLASAAASSRAGIAPPLAGVSPGSSSYSPGSSSYSPGSSSYSPGSSSY 382  
OY 383 AAGGAGAGTGHCHNLOAMSLYAAERGHLOAGPAGGSAVDLPDYSLPVTSSSSS 442  
Db 383 AAGGAGAGTGHCHNLOAMSLYAAERGHLOAGPAGGSAVDLPDYSLPVTSSSSS 442  
OY 443 SLSHGGGGGGGGGGGGHHPAAGHGRITSMYLNQAGGDLGHILASAAAAAAGYGGQ 502  
Db 443 SLSHGGGGGGGGGGGGHHPAAGHGRITSMYLNQAGGDLGHILASAAAAAAGYGGQ 502  
OY 503 NFHSVREMFESQRIGLNNSPVNGSSCOMAFSSQSLYRTSGAFVYDCKF 553  
Db 503 NFHSVREMFESQRIGLNNSPVNGSSCOMAFSSQSLYRTSGAFVYDCKF 553

RESULT 2  
US-09-083-351-5  
; Sequence 5, Application US/09083351  
; Patent No. 6087107  
GENERAL INFORMATION:  
APPLICANT: Sheffield, Val C.  
APPLICANT: Alward, Wallace L.M.  
APPLICANT: Stone, Edwin M.  
APPLICANT: Nishimura, Darryl  
APPLICANT: Patil, Shiva  
TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,351  
FILING DATE: 22-MAY-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: UIA-029.02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 106 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-083-351-5

Query Match 19.2%; Score 106; DB 3; Length 106;  
Best Local Similarity 100.0%; Pred. No. 2.3e-87;

Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 73 PKDMKPPRYSTALITMAIONAPDKITLNGIYOFIMDRFPFYRDNKGOMNSIRNLST 132  
Db 1 PKDMKPPRYSTALITMAIONAPDKITLNGIYOFIMDRFPFYRDNKGOMNSIRNLST 60  
OY 133 NECFVKVPRDCKKRGKGSYWTLPDPSYNMFENGSLRRRRFRKKD 178  
Db 61 NECFVKVPRDCKKRGKGSYWTLPDPSYNMFENGSLRRRRFRKKD 106

RESULT 3  
US-09-083-351-6  
; Sequence 6, Application US/09083351  
; Patent No. 6087107  
GENERAL INFORMATION:  
APPLICANT: Sheffield, Val C.  
APPLICANT: Alward, Wallace L.M.  
APPLICANT: Stone, Edwin M.  
APPLICANT: Nishimura, Darryl  
APPLICANT: Patil, Shiva  
TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,351  
FILING DATE: 22-MAY-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: UIA-029.02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 106 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-083-351-6

Query Match 11.0%; Score 61; DB 3; Length 106;  
Best Local Similarity 100.0%; Pred. No. 3.6e-47;  
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 118 NKGOMNSIRHNLSLNECFVKVPRDCKKRGKGSYWTLPDPSYNMFENGSLRRRRFRKK 177  
Db 46 NKGOMNSIRHNLSLNECFVKVPRDCKKRGKGSYWTLPDPSYNMFENGSLRRRRFRKK 105  
OY 178 D 178  
Db 106 D 106  
RESULT 4  
US-09-083-351-7

Sequence 7, Application US/09083351  
Patent No. 6087107  
GENERAL INFORMATION:  
APPLICANT: Sheffield, Val C.  
APPLICANT: Alward, Wallace L.M.  
APPLICANT: Stone, Edwin M.  
APPLICANT: Nishimura, Darryl  
APPLICANT: Patil, Shiva  
TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
TITLE OF INVENTION: TRANSCRIPTION FACTOR  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,351  
FILING DATE: 22-MAY-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: UIA-029.02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 106 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-083-351-7

Query Match 4.2%; Score 23; DB 3; Length 106;  
Best Local Similarity 100.0%; Pred. No. 3.1e-13;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GWNSTIRNLSINECFVAVPRDD 143  
|||||  
DB 49 GWNSTIRNLSINECFVAVPRDD 71

RESULT 5  
US-09-083-351-8  
Sequence 8, Application US/09083351  
Patent No. 6087107  
GENERAL INFORMATION:  
APPLICANT: Sheffield, Val C.  
APPLICANT: Alward, Wallace L.M.  
APPLICANT: Stone, Edwin M.  
APPLICANT: Nishimura, Darryl  
APPLICANT: Patil, Shiva  
TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
TITLE OF INVENTION: TRANSCRIPTION FACTOR  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA

COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,351  
FILING DATE: 22-MAY-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: UIA-029.02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 106 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-083-351-8

Query Match 2.9%; Score 16; DB 3; Length 106;  
Best Local Similarity 100.0%; Pred. No. 5.6e-07;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 TLNGTVOFIMDRPEFY 115  
|||||  
DB 28 TLNGTVOFIMDRPEFY 43

RESULT 6  
US-09-083-351-16  
Sequence 16, Application US/09083351  
Patent No. 6087107  
GENERAL INFORMATION:  
APPLICANT: Sheffield, Val C.  
APPLICANT: Alward, Wallace L.M.  
APPLICANT: Stone, Edwin M.  
APPLICANT: Nishimura, Darryl  
APPLICANT: Patil, Shiva  
TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
TITLE OF INVENTION: TRANSCRIPTION FACTOR  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,351  
FILING DATE: 22-MAY-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: UIA-029.02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000

TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 106 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-083-351-16

Query Match  
Best Local Similarity 2.9%; Score 16; DB 3; Length 106;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 NKOGWONSIRHNLSLN 133  
DB 46 NKOGWONSIRHNLSLN 61

RESULT 7  
US-09-083-351-17  
Sequence 17, Application US/09083351  
Patent No. 6087107  
GENERAL INFORMATION:  
APPLICANT: Sheffield, Val C.  
APPLICANT: Alward, Wallace L.M.  
APPLICANT: Stone, Edwin M.  
APPLICANT: Nishimura, Dairryl  
APPLICANT: Patil, Shiva  
TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY, HOG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,351  
FILING DATE: 22-MAY-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: UIA-029.02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 106 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-083-351-17

Query Match  
Best Local Similarity 2.9%; Score 16; DB 3; Length 106;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 NKOGWONSIRHNLSLN 133  
DB 46 NKOGWONSIRHNLSLN 61

DB 46 NKOGWONSIRHNLSLN 61

RESULT 8  
US-07-882-292-2  
Sequence 2, Application US/07882292  
Patent No. 5324638  
GENERAL INFORMATION:  
APPLICANT: Tao, Wufan  
APPLICANT: Lai, Eseng  
TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC ACIDS  
TITLE OF INVENTION: ENCODING SAME AND USES THEREOF  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John P. White  
STREET: c/o Cooper and Dunham, 30 Rockefeller  
STREET: Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/882,292  
FILING DATE: 19920513  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 41472  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-977-9550  
TELEFAX: 212-664-0525  
TELLEX: 422523 COOP UI  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 480 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-882-292-2

Query Match  
Best Local Similarity 2.9%; Score 16; DB 1; Length 480;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 NKOGWONSIRHNLSLN 133  
DB 212 NKOGWONSIRHNLSLN 227

RESULT 9  
US-08-331-644-2  
Sequence 2, Application US/08331644  
Patent No. 5976872  
GENERAL INFORMATION:  
APPLICANT: Tao, Wufan  
APPLICANT: Lai, Eseng  
TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC  
TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/331,644  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/882,292  
FILING DATE: 13-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 41472-A-PCT-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-391-0525  
TELEFAX: 212-278-0400  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 480 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-331-644-2

Query Match 2.9%; Score 16; DB 2; Length 480;  
Best Local Similarity 100.0%; Pred. No. 2.3e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 NKQGMONSIRHNSLN 133  
DB 212 NKQGMONSIRHNSLN 227

RESULT 10  
PCT-US93-04102-2  
Sequence 2, Application PC/TUS9304102  
GENERAL INFORMATION:  
APPLICANT: Tao, Wufan  
APPLICANT: Lai, Eseng  
TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC  
TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John P. White  
STREET: c/o Cooper and Dunham, 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/04102  
FILING DATE: 19930430  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/882,292  
FILING DATE: 13-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 41472A-PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-664-0525  
TELEFAX: 212-664-0525  
TELEX: 422523 COOP UI

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 480 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US93-04102-2

Query Match 2.9%; Score 16; DB 4; Length 480;  
Best Local Similarity 100.0%; Pred. No. 2.3e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 NKQGMONSIRHNSLN 133  
DB 212 NKQGMONSIRHNSLN 227

RESULT 11  
US-09-083-351-12  
Sequence 12, Application US/09083351  
Patent No. 6087107  
GENERAL INFORMATION:  
APPLICANT: Sheffield, Val C.  
APPLICANT: Alward, Wallace L.M.  
APPLICANT: Stone, Edwin M.  
APPLICANT: Nishimura, Darryl  
APPLICANT: Patil, Shiva  
TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY, HONG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,351  
FILING DATE: 22-MAY-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: UIA-029.02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 106 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-083-351-12

Query Match 2.7%; Score 15; DB 3; Length 106;  
Best Local Similarity 100.0%; Pred. No. 4.3e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 VKPPSYALITMAI 91  
DB 5 VKPPSYALITMAI 19

RESULT 12  
US-09-083-351-13  
; Sequence 13, Application US/09083351  
; Patent No. 6087107  
; GENERAL INFORMATION:  
; APPLICANT: Sheffield, Val C.  
; APPLICANT: Alward, Wallace L.M.  
; APPLICANT: Stone, Edwin M.  
; APPLICANT: Nishimura, Darryl  
; APPLICANT: Patil, Shiva  
; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
; TITLE OF INVENTION: TRANSCRIPTION FACTOR  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
; STREET: One Post Office Square  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109-2170  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/083,351  
; FILING DATE: 22-MAY-1998  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arnold, Beth E.  
; REGISTRATION NUMBER: 35,430  
; REFERENCE/DOCKET NUMBER: UIA-029.02  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-832-1000  
; TELEFAX: 617-832-7000  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 106 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-083-351-13

Query Match 2.7%; Score 15; DB 3; Length 106;  
Best Local Similarity 100.0%; Pred. No. 4.3e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 VKPPSYXIALITMAI 91  
Db 5 VKPPSYXIALITMAI 19

RESULT 13  
US-09-083-351-19  
; Sequence 19, Application US/09083351  
; Patent No. 6087107  
; GENERAL INFORMATION:  
; APPLICANT: Sheffield, Val C.  
; APPLICANT: Alward, Wallace L.M.  
; APPLICANT: Stone, Edwin M.  
; APPLICANT: Nishimura, Darryl  
; APPLICANT: Patil, Shiva  
; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
; TITLE OF INVENTION: TRANSCRIPTION FACTOR  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP

STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,351  
FILING DATE: 22-MAY-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: UIA-029.02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 76 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-083-351-19

Query Match 2.4%; Score 13; DB 3; Length 76;  
Best Local Similarity 100.0%; Pred. No. 0.00019;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GWNOSIRHNLISLN 133  
Db 49 GWNOSIRHNLISLN 61

RESULT 14  
US-07-906-930E-4  
; Sequence 4, Application US/07906930E  
; Patent No. 5534631  
; GENERAL INFORMATION:  
; APPLICANT: Gaynor, Richard B.  
; APPLICANT: Nirula, Ajay  
; APPLICANT: Li, Ching  
; TITLE OF INVENTION: DNA ENCODING THE INTERLEUKIN BINDING  
; TITLE OF INVENTION: FACTOR (ILF)  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P. O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: USA  
; ZIP: 77210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/906,930E  
; FILING DATE: 30-JUN-1992  
; CLASSIFICATION: 336  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sertlich, Gary J.  
; REGISTRATION NUMBER: 34,430  
; REFERENCE/DOCKET NUMBER: UTSD:262/SER  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 512-418-3000

TELEFAX: 512-474-7577  
 TELEX: NOT APPLICABLE  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 98 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-07-906-930E-4

Query Match 2.4%; Score 13; DB 1; Length 98;  
 Best Local Similarity 100.0%; Pred. No. 0.00025;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 121 GWNSIRHNLN 133  
 |||  
 Db 46 GWNSIRHNLN 58

RESULT 15  
 US-09-083-351-11  
 ; Sequence 11, Application US/09083351  
 ; Patent No. 6087107  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sheffield, Val C.  
 ; APPLICANT: Alward, Wallace L.M.  
 ; APPLICANT: Stone, Edwin M.  
 ; APPLICANT: Nishimura, Darryl  
 ; APPLICANT: Patil, Shiva  
 ; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
 ; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
 ; TITLE OF INVENTION: TRANSCRIPTION FACTOR  
 ; NUMBER OF SEQUENCES: 22  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: FOLEY, HONG & ELIOT LLP  
 ; STREET: One Post Office Square  
 ; CITY: Boston  
 ; STATE: MA  
 ; COUNTRY: USA  
 ; ZIP: 02109-2170  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC Compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/083,351  
 ; FILING DATE: 22-MAY-1998  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Arnold, Beth E.  
 ; REGISTRATION NUMBER: 35,430  
 ; REFERENCE/DOCKET NUMBER: UIA-029.02  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617-832-1000  
 ; TELEFAX: 617-832-7000  
 ; INFORMATION FOR SEQ ID NO: 11:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 106 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS:  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-09-083-351-11

Query Match 2.4%; Score 13; DB 3; Length 106;  
 Best Local Similarity 100.0%; Pred. No. 0.00027;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 121 GWNSIRHNLN 133

Db 49 GWNSIRHNLN 61  
 |||

Search completed: February 6, 2001, 22:53:53  
 Job time: 8164 sec





Wed Feb 7 14:05:55 2001

us-09-292-862-1.lst

Page 1

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 6, 2001, 18:03:13 ; Search time 966.16 Seconds  
(without alignments)  
12054.364 Million cell updates/sec

Title: US-09-292-862-1  
Perfect score: 1662  
Sequence: 1 atgcagcgcgcctactcgtl.....acgactgttagcaagtlttga 1662

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 80.0

Searched: 7991742 seqs, 3503743858 residues

Word size : 0 15983484

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_estba:\*  
42: em\_estfm:\*  
43: em\_esthum1:\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estov1:\*  
68: em\_estov2:\*  
69: em\_estpl1:\*  
70: em\_estpl2:\*  
71: em\_estpl3:\*  
72: em\_estpl4:\*  
73: em\_estpl5:\*  
74: em\_estro1:\*  
75: em\_estro2:\*  
76: em\_estro3:\*  
77: em\_estro4:\*  
78: em\_estro5:\*  
79: em\_estro6:\*  
80: em\_estro7:\*  
81: em\_estro8:\*  
82: em\_estro9:\*  
83: em\_estro10:\*  
84: em\_estro11:\*  
85: em\_estro12:\*  
86: em\_estro13:\*  
87: gb\_esta1:\*  
88: gb\_esta2:\*  
89: gb\_esta3:\*  
90: gb\_esta4:\*  
91: gb\_esta5:\*  
92: gb\_esta6:\*  
93: gb\_esta7:\*  
94: gb\_esta8:\*  
95: gb\_esta9:\*  
96: gb\_esta10:\*  
97: gb\_esta11:\*  
98: gb\_esta12:\*  
99: gb\_esta13:\*  
100: gb\_esta14:\*  
101: gb\_esta15:\*  
102: gb\_esta16:\*  
103: gb\_esta17:\*  
104: gb\_esta18:\*  
105: gb\_esta19:\*  
106: gb\_esta20:\*  
107: gb\_esta21:\*  
108: gb\_esta22:\*  
109: gb\_esta23:\*  
110: gb\_esta24:\*  
111: gb\_esta25:\*  
112: em\_esthum21:\*  
113: em\_esthum22:\*  
114: em\_esthum23:\*  
115: em\_estom1:\*  
116: em\_estom2:\*

117: em\_estp16:\*  
 118: em\_estp17:\*  
 119: em\_estp18:\*  
 120: em\_estp19:\*  
 121: em\_estp20:\*  
 122: em\_estp21:\*  
 123: em\_estp22:\*  
 124: em\_estp23:\*  
 125: em\_estp24:\*  
 126: em\_estp25:\*  
 127: em\_estp26:\*  
 128: em\_estp27:\*  
 129: em\_estp28:\*  
 130: em\_estp29:\*  
 131: em\_estp30:\*  
 132: em\_estp31:\*  
 133: em\_estp32:\*  
 134: em\_estp33:\*  
 135: em\_estp34:\*  
 136: em\_estp35:\*  
 137: em\_estp36:\*  
 138: em\_estp37:\*  
 139: em\_estp38:\*  
 140: em\_estp39:\*  
 141: em\_estp40:\*  
 142: em\_estp41:\*  
 143: em\_estp42:\*  
 144: em\_estp43:\*  
 145: em\_estp44:\*  
 146: em\_estp45:\*  
 147: em\_estp46:\*  
 148: em\_estp47:\*  
 149: em\_estp48:\*  
 150: em\_estp49:\*  
 151: em\_estp50:\*  
 152: em\_estp51:\*  
 153: em\_estp52:\*  
 154: em\_estp53:\*  
 155: em\_estp54:\*  
 156: em\_estp55:\*  
 157: em\_estp56:\*  
 158: em\_estp57:\*  
 159: em\_estp58:\*  
 160: em\_estp59:\*  
 161: em\_estp60:\*  
 162: em\_estp61:\*  
 163: em\_estp62:\*  
 164: em\_estp63:\*  
 165: em\_estp64:\*  
 166: em\_estp65:\*  
 167: em\_estp66:\*  
 168: em\_estp67:\*  
 169: em\_estp68:\*  
 170: em\_estp69:\*  
 171: em\_estp70:\*  
 172: em\_estp71:\*  
 173: em\_estp72:\*  
 174: em\_estp73:\*  
 175: em\_estp74:\*  
 176: em\_estp75:\*  
 177: em\_estp76:\*  
 178: em\_estp77:\*  
 179: em\_estp78:\*  
 180: em\_estp79:\*  
 181: em\_estp80:\*  
 182: em\_estp81:\*  
 183: em\_estp82:\*  
 184: em\_estp83:\*  
 185: em\_estp84:\*  
 186: em\_estp85:\*  
 187: em\_estp86:\*  
 188: em\_estp87:\*  
 189: em\_estp88:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	305	18.4	305	23	A1659865	A1659865 tu18h10.x
2	263	15.8	555	24	A1761460	A1761460 w65h10.x
3	249	15.0	305	25	A1817619	A1817619 w39f102.x
4	231	13.9	507	19	A1333220	A1333220 q905g08.x
5	186	11.2	504	25	A1819982	A1819982 w760a05.x
6	185	11.1	259	90	AA451944	AA451944 UI-H-B13-
7	185	11.1	292	90	AA451610	AA451610 UI-H-B13-
8	169	10.2	439	1	AA022618	AA022618 ze71a01.s
9	163	9.8	223	25	A1804844	A1804844 t955a09.x
10	156	9.4	180	24	A1763387	A1763387 w155h12.x
11	155	9.3	579	87	AW195359	AW195359 x037d07.x
12	150	9.0	242	88	AA263103	AA263103 x035h09.x
13	138	8.3	138	148	AF190834	AF190834 AF190834
14	135	8.1	486	110	BE644653	BE644653 7e68f05.x
15	112	6.7	436	20	A1453830	A1453830 t347e05.x
16	83	5.0	131	23	A1701617	A1701617 t298b04.x
17	74	4.5	232	21	A1508825	A1508825 vc30a07.y
18	74	4.5	440	4	AA276025	AA276025 vc30a07.y
19	74	4.5	489	20	A1464122	A1464122 vc30a07.x
20	68	4.1	327	88	AA271272	AA271272 xs08g07.x
21	63	3.8	429	16	A1111819	A1111819 UI-R-YO-m
22	55	3.3	442	24	A1713280	A1713280 UI-R-AB1-
23	55	3.3	460	16	A1112837	A1112837 UI-R-E1-f
24	54	3.2	122	147	W81946	W81946 me94f07.r1
25	53	3.2	196	27	A1935186	A1935186 wpl14h02.x
26	49	2.9	147	95	AW793237	AW793237 CM4-UM002.x
27	49	2.9	175	136	BE858880	BE858880 7948e08.x
28	47	2.8	466	18	A1255203	A1255203 u132f04.y
29	47	2.8	623	10	AA673797	AA673797 vu08f03.x
30	41	2.5	415	91	AA495202	AA495202 UI-M-BH3-
31	41	2.5	1009	190	CNS01SIF	AL165264 Tetraodon
32	35	2.1	363	93	AM653362	AM653362 Tetraodon
33	35	2.1	471	192	CNS030DB	AL164622 Tetraodon
34	35	2.1	951	190	CNS01SOL	AA998341 UI-R-C0-1
35	33	2.0	417	14	AA998341	AA998341 UI-R-C0-1
36	33	2.0	535	39	AM025430	AM025430 wu74b06.x
37	32	1.9	82	14	AA960591	AA960591 wv64c01.s
38	32	1.9	215	24	A1713590	A1713590 UI-R-AG1-
39	32	1.9	235	20	A1463604	A1463604 wv64c01.x
40	32	1.9	330	20	A1415347	A1415347 md53e12.x
41	32	1.9	330	88	AA253193	AA253193 UI-R-BU0-
42	32	1.9	338	25	A1844747	A1844747 UI-M-ALI-
43	32	1.9	472	20	A1427433	A1427433 m04f01.x
44	32	1.9	484	20	A1706080	A1706080 UI-R-AB1-
45	32	1.9	485	24	A1706734	A1706734 UI-R-AA1-

## ALIGNMENTS

RESULT 1  
 LOCUS A1659865/C  
 DEFINITION tu18h10.x1 NCI-CCAP\_P128 Homo sapiens CDNA clone IMAGE:2251459 3', similar to TR:093440 093440 WINGED-HELIX TRANSCRIPTION FACTOR.  
 ACCESSION A1659865  
 VERSION A1659865.1 GI:4763435  
 KEYWORDS EST.



1

1

Db	1	GGGGCCGACGAGCTACCCGGGCGACGACGAGAACTCCACTCGGTCGGGAGAGTTTCGAG	50
QY	1537	tcacagagagtcgcgttgaacaactctccagtgaacgggaatatgaacttcaaatggcc	1596
Db	61	TCACGAGAGATCGGCTTGAACAACACTCTCCAGTGAACGGGAATGAGTGTCAATATGCC	120
QY	1597	ttccctccagcagagctctctgtacccagctccggagcttcctgcttaacgactgaacgaag	1656
Db	121	TTCCCTTCACACCACTCTCTGTACCGCAGAGTCGGAGCTTTGTCTACGACTGTACGCAAG	180
QY	1657	ttttga 1662	
Db	181	TTTTGA 186	
RESULT	6		
AM451944			
DEFINITION	259 bp mRNA	EST	17-FEB-2000
LOCUS	UI-H-B13-alt.-g.-08-0-UI.s1 NCI-CGAP-Sub5 Homo sapiens cDNA clone		
ACCESSION	IMAGE:3068727 3', mRNA sequence.		
VERSION	AM451944		
SOURCE	AM451944.1 GI:692720		
KEYWORDS	EST.		
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
JOURNAL	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)		
	Contact: Robert Strausberg, Ph.D.		
	Tel: (301) 496-1550		
	Email: Robert.Strausberg@nih.gov		
	oligo-dr track not found, Not I site shown in beginning of sequence		
	is likely internal to the message. cDNA library preparation: M.B.		
	Soares Lab Clone distribution: NCI-CGAP clone distribution		
	Information can be found through the I.M.A.G.E. Consortium/LLNL at:		
	www.bio.llnl.gov/dbfp/image/image.html		
	Seq primer: M13 Forward		
	POLYA=NO.		
FEATURES	Location/Qualifiers		
SOURCE	1. 259		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:3068727"		
	/clone_lib="NCI-CGAP-Sub5"		
	/lab_host="DH10B (Life Technologies)"		
	/note="Vector: pRT3D-Pac (Pharmacia) with a modified		
	polylinker. Site.1: Not I; Site.2: Eco RI; NCI-CGAP-Sub5		
	is a subtracted library derived from NCI-CGAP-Sub4. The		
	NCI-CGAP-Sub5 library had 3 million recombinants. A		
	single-stranded DNA preparation of NCI-CGAP-Sub4 was used		
	as a tracer in a subtractive hybridization with a driver		
	comprising: the IMAGE pool (NCI-CGAP-K13 pool 1 LLM		
	3334-3337, 3682-3683, 3798-3803 (IMAGE Clones		
	1322376-1323911, 1456008-1456775, 1500552-1502855);		
	NCI-CGAP Kids pool 1 LLM 3338-3342, 3722-3725, 3776-3778		
	(IMAGE Clones 1323912-1325831, 1471368-1472903,		
	1492104-1493255); NCI-CGAP LWS pool 1 LLM 3575-3582,		
	3851-3854 (IMAGE Clones 1414920-1417991, 1520904-152243		
	); NCI-CGAP-GC4 pool 1 LLM 3164-3167, 3716-3720,		
	3733-3735 (IMAGE Clones 1257096-1258631, 1469064-1470983		
	1475592-1476743); NCI-CGAP-P22 pool 1 LLM 2457-2459,		
	2758-2759, 3062-3068 (IMAGE Clones 98508-986759		
	1101192-1101959, 121928-1220615); NCI-CGAP-Co10 pool 1		
	LLM 2644-2653, 2871-2872 (IMAGE Clones 1057416-1061255,		
	1143584-1145351). (10% of the driver population), plus a		
	pool of 3,840 arrayed clones from NCI-CGAP-Sub1 (IMAGE		
	Clones 2708616-2710535) and NCI-CGAP-Sub2 (IMAGE		
	Clones 2710536-2712455) (10% of the driver population		





TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Ian Kirsch, M.D., Michael R. Emmert-Buck, M.D.,  
Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/bbrp/image/image.html  
Insert length: 268 Std Error: 0.00  
Seq primer: -40UP from Gibco.  
Location/Qualifiers

## FEATURES

source

1. 180  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="2394215"  
/clone\_lib="NCI-CGAP\_Col16"  
/tissue\_type="colon tumor, RER+"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI-CGAP\_Col16 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneIDs 1057416-1061255, and 1144584-1145351).  
Subtraction by Bento Soares and M. Fatima Bonaldo. "  
BASE COUNT 45 a 51 c 42 g 42 t  
ORIGIN

Query Match  
Best Local Similarity 9.4%; Score 156; DB 24; Length 180;  
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1507 aactccactcgtgctgagatggttcagtcacagagatcgtcttaaacacttcca  
Db 1 AACTTCACACGCGGTGGCGGAGATGTCAGTCACAGAGATGCGCTTGACACACTCTCCA 60  
OY 1567 gtgaacgggaatagtagctgcaaatgagccttcacagccagctctctgaccgacg 1626  
Db 61 GTGAACGGGAATAGTAGCTGTCAAAATGGCTTCCTCCAGCCAGTCTGTGTACCCACG 120  
OY 1627 tccggagcttcgctcagactgacgttagcaagtttga 1662  
Db 121 TCCGAGCTTTCGTCTACAGCTGTAGCAAGTTTGA 156

## RESULT 11

AM195359

## LOCUS

AM195359 579 bp mRNA

EST 29-NOV-1999

3'

2'

1'

similar to SW:MFHL\_HUMAN Q99558 MENSENHYME FORK HEAD PROTEIN 1 ;

MRNA sequence.

AM195359

EST

GI:6474435

EST.

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

Tumor Gene Index

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/bbrp/image/image.html

## FEATURES

source

1. 579  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="2635885"  
/clone\_lib="NCI-CGAP\_Kid11"  
/lab\_host="DH10B"  
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI-CGAP\_Kid11 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneIDs 1322376-1323911, 1456007-1456775, and  
1500552-1502855). Subtraction by Bento Soares and M.  
Fatima Bonaldo. "  
BASE COUNT 180 a 161 c 92 g 146 t  
ORIGIN

Query Match  
Best Local Similarity 9.3%; Score 155; DB 87; Length 579;  
Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1508 aactccactcgtgctgagatggttcagtcacagagatcgtcttaaacacttcca  
Db 1 AACTTCACACGCGGTGGCGGAGATGTCAGTCACAGAGATGCGCTTGACACACTCTCCA 60  
OY 1568 tgaacgggaatagtagctgcaaatgagccttcacagccagctctctgaccgacg 1627  
Db 61 TGAACGGGAATAGTAGCTGTCAAAATGGCTTCCTCCAGCCAGTCTGTGTACCCACG 120  
OY 1628 ccggaacttcgctcagactgacgttagcaagtttga 1662  
Db 121 CCGAGCTTTCGTCTACAGCTGTAGCAAGTTTGA 155

## RESULT 12

AM263103

## LOCUS

AM263103 242 bp mRNA

EST 28-DEC-1999

3'

2'

1'

similar to SW:MFHL\_HUMAN Q99558 MENSENHYME FORK HEAD PROTEIN 1 ;

MRNA sequence.

AM263103

EST

GI:6639919

EST.

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

Tumor Gene Index

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: similarity on wrong strand  
Possible reversed clone: polyt not found  
Seq primer: -40UP from Gibco  
High quality sequence stop: 461.  
Location/Qualifiers





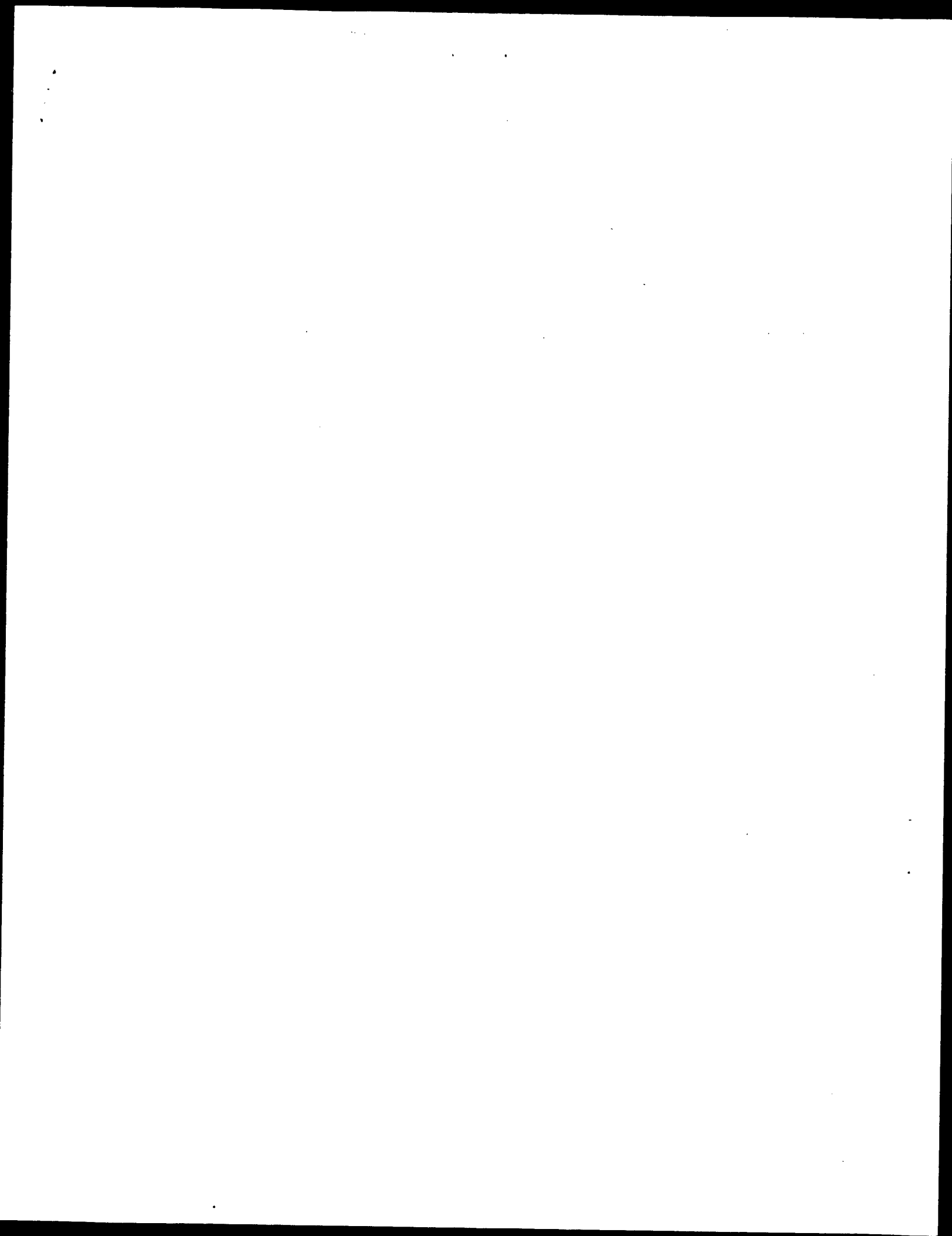


Wed Feb 7 14:05:55 2001

us-09-292-862-1.rst

---

Page 11



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 6, 2001, 18:24:39 : Search time 51.77 Seconds  
(without alignments)  
5173.815 Million cell updates/sec

Title: US-09-292-862-1

Perfect score: 1662  
Sequence: 1 atgcagcgccgctactcgcgt.....acgactgtagcaatttga 1662

Scoring table: OLIGO\_NUC  
Gapop 60.0, Gapext 60.0

Searched: 280836 segs, 80580151 residues

Word size: 0

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/PCITUS.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1217	73.2	3946	3	US-09-083-351-1
2	1214	73.0	1659	3	US-09-083-351-3
3	47	2.8	1155	2	US-08-331-644-4
4	47	2.8	1155	4	PCT-US93-04102-4
5	28	1.7	1860	2	US-08-331-644-3
6	28	1.7	1860	4	PCT-US93-04102-3
7	27	1.6	640	2	US-08-835-099A-16
8	27	1.6	640	2	US-09-157-349-16
9	27	1.6	804	2	US-08-835-099A-10
10	27	1.6	804	3	US-09-157-349-10
11	25	1.5	3330	1	US-08-149-103-1
12	25	1.5	3330	1	US-08-451-883-1
13	25	1.5	3656	1	US-08-393-734-1
14	25	1.5	8438	1	US-07-945-283-1
15	25	1.5	9592	1	US-08-393-734-3
16	25	1.5	25	1	US-08-374-144-3
17	24	1.4	25	1	US-08-775-164-3
18	24	1.4	25	2	US-08-775-609-3
19	24	1.4	25	2	US-08-775-607-3
20	24	1.4	30	4	PCT-US93-06828-3
21	24	1.4	30	4	US-08-068-747-2
22	24	1.4	33	1	US-08-068-747-7
23	24	1.4	69	3	US-08-480-173A-31
24	24	1.4	69	3	US-08-480-173A-32
25	24	1.4	69	3	US-08-484-408A-31
26	24	1.4	69	3	US-08-484-408A-32
27	24	1.4	103	3	US-08-480-173A-12
28	24	1.4	103	3	US-08-484-408A-12

29	24	1.4	457	2	US-08-332-765A-7	Sequence 7, Appl
30	24	1.4	519	2	US-08-581-528A-5	Sequence 5, Appl
31	24	1.4	519	4	PCT-US94-07799-5	Sequence 1, Appl
32	24	1.4	633	1	US-08-234-783-1	Sequence 1, Appl
33	24	1.4	633	1	US-08-456-907-1	Sequence 1, Appl
34	24	1.4	633	4	PCT-US95-05523-1	Sequence 29, Appl
35	24	1.4	1203	1	US-08-362-670B-29	Sequence 29, Appl
36	24	1.4	1203	3	US-08-333-576C-29	Sequence 29, Appl
37	24	1.4	1203	4	PCT-US94-14030A-29	Sequence 29, Appl
38	24	1.4	1393	2	US-08-765-875-1	Sequence 1, Appl
39	24	1.4	1393	3	US-08-795-671-1	Sequence 1, Appl
40	24	1.4	1769	2	US-08-765-268A-1	Sequence 1, Appl
41	24	1.4	1809	4	US-08-455-001-1	Sequence 1, Appl
42	24	1.4	1809	4	PCT-US95-11869-1	Sequence 1, Appl
43	24	1.4	2953	2	US-08-859-201-1	Sequence 1, Appl
44	24	1.4	3090	4	PCT-US93-06251-7	Sequence 7, Appl
45	24	1.4	3765	3	US-07-705-490-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-09-083-351-1  
Sequence 1, Application US/09083351  
Patent No. 6087107  
GENERAL INFORMATION:  
APPLICANT: Sheffield, Val C.  
APPLICANT: Alward, Wallace L.M.  
APPLICANT: Stone, Edwin M.  
APPLICANT: Nishimura, Darryl  
APPLICANT: Patil, Shiva  
TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,351  
FILING DATE: 22-MAY-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: UIA-029.02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-7000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3946 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 475..2133  
US-09-083-351-1  
Query Match 73.2%; Score 1217; DB 3; Length 3946;

Best Local Similarity 99.6%; Pred. No. 0;  
Matches 1657; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

```

QY 1 atgcagcgccgctactccgcttccagccccaactccctggaagtgtgacctactcgc 60
Db 475 ATGAGGCGCGCTACTCCGTGTCAGCCCACTCCCTGGAGTGTGCTTACTCCGCG 534
QY 61 ggcagcaagactactaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 120
Db 535 GCGGAGCAAGCTACTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 594
QY 121 ccggccccctgagcgtgtactcgcacctgcgcgcgcgcgcgcgcgcgcgcgc 180
Db 595 CCGGCCCCCATGAGGTGCTACTCGCACCTCGCGCACGCCGAGCACTACCGGGCGCG 654
QY 181 gcccgcctactaggccctactaccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240
Db 655 GCGCGCGCTTACGGGCGCTTACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 714
QY 241 tatagctacatcgctcatalcaccatgacatgcagaaagcccggaagaataacc 300
Db 715 TATAGCTCATGCGCGCTCATCACCATGCGCATGCGCGCGCGCGCGCGCGCGCG 774
QY 301 ctgagcgcatctaccatctcattcattgacgcgcctccctctacccggaacaag 360
Db 775 CTGAGCGCATCTACCAATGTTATGATGACCGCTTCCCTTCTACCGGAGCAACA 834
QY 361 ggcctggcaaacacatccgcacacacctcgcctcagcagagtgcttcgtcaag 420
Db 835 GCGTGGCAAGACAGATCCGCCACACCTTCGCTCAAGAGTCTTCTGTCAGAGT 894
QY 421 cgcgaagcaagaaagccggcgaagcagctggaagcgcgcgcgcgcgcgcgcgc 480
Db 895 CGGACGCAAGAAAGCCGGGCAAGCGAGCTGAGAGCGTGGAGCCGCGAGCTCTA 954
QY 481 atgttcagaaagcagcttccctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
Db 955 ATGTTGAGAAAGCGCTTCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1014
QY 541 aaggaaggaaggaaggaagcagctgcaacctcaagagagcgcgcgcgcgcgc 599
Db 1015 AAGGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1074
QY 600 ccccgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 659
Db 1075 CCCCCCGCCC-GCGCGCGCGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1133
QY 660 cgttgcgcatccaggaatcaagacgcgaagcgttacgttgcctcgcgcgcgc 719
Db 1134 CGTGGCATCTCAAGACATCAAGACCGAGAAAGGTAGTGCCTCGCGCGCGCGCG 1193
QY 720 gtcccgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 779
Db 1194 GTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1253
QY 780 cgaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 839
Db 1254 CGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1313
QY 840 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 899
Db 1314 GCGGCGCGCTAGCGGTGAGCGGTGCGGATTCGCGCGCGCGCGCGCGCGCGCG 1373
QY 900 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 959
Db 1374 GCGCGCGCACCTATGCGAGGCTTCAAGCTTGAGCAACATCATGAGTGTGCGG 1433
QY 960 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1019
Db 1434 GCGCGAGAGCGCGCGCGCGAGAGTCAAGCTTCTGCGCTGCGCGCGCGCGCG 1493
QY 1020 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1079

```

```

Db 1494 GCGCGCGGAGATCGCACCCCGCTGCGCTGCGGCGCTACTCGCGCGCGAGAGCTCCT 1553
QY 1080 ctacagctccccctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1139
Db 1554 CTACAGCTCCCCCTGCGAGCGCAAGACTTCTCAAGCGCGGAGCTCGCGCGCGCGCG 1613
QY 1140 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1199
Db 1614 GCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1673
QY 1200 gacgtctagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1259
Db 1674 GAGCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1733
QY 1260 cgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1319
Db 1734 CGGCTGCGCGGTGAGCAACCCCTGCGCGAGCTTCTTGTCTCGGTCAAGACAGCAG 1793
QY 1320 ctgcgtccctgagtcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1379
Db 1794 CTGCTGCTCCTGATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1853
QY 1380 ccaacacccctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1439
Db 1854 CCACACACCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1913
QY 1440 agacgtggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1499
Db 1914 AGACTGGGCGCACTTGGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1973
QY 1500 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1559
Db 1974 GCGCGAAGCTTCACTCGCTCGGAGAGTGTTCAGATCAGAGAGATCGGCTTGA 2033
QY 1560 ctctcagtgaaacgcgaatagctgtcaaaagcgtccctccacgcgcgcgcgc 1619
Db 2034 CTCTCAGTGAAGCGGAATAGTGTGCTCAATGAGCTTCTTCCCTTACAGCACTCT 2093
QY 1620 ccgcacgtccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1662
Db 2094 CCGCACGTCCGAGGCTTCTGCTACAGCTGAGCAAGTTTGA 2136

RESULT 2
US-09-083-351-3
; Sequence 3, Application US/09083351
; Patent No. 6087107
; GENERAL INFORMATION:
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; APPLICANT: Stone, Edwin M.
; APPLICANT: Nishimura, Darryl
; APPLICANT: Patil, Shiva
; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR
; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & FLIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,351
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514

```



us-09-292-862-1.rnj

RESULT 4  
PCT-US93-04102-4  
Sequence 4, Application PC/TUS9304102  
GENERAL INFORMATION:  
APPLICANT: Tao, Wufan  
APPLICANT: Lai, Eseng  
TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NOCUTIC  
TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John P. White  
STREET: c/o Cooper and Dunham, 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/04102  
FILING DATE: 19930430  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/882,292  
FILING DATE: 13-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.

```

RESULT 5
US-08-331-644-3
; Sequence 3, Application US/08331644
; Patent No. 5976872
GENERAL INFORMATION:
APPLICANT: Tao, Wufan
APPLICANT: Lai, Eseng
TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/331,644
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/882,292
FILING DATE: 13-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41472-A-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1860 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-331-644-3
Query Match 1.7%; Score 28; DB 2; Length 1860.
Best Local Similarity 100.0%; Pctd. No. 0.0048;

```





Query Match	1.6%	Score 27;	DB 3;	Length 804;
Best Local Similarity	100.0%	Pred. No. 0.014;		
Matches	27;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;

OY 1342 ggcggcgccgagcgccgagcg 1368  
|||||  
DB 28 GCGCGCGCGCGCGCGCGCGCGG 54

RESULT 11  
US-08-149-103-1  
Sequence 1, Application US/08149103  
Patent No. 5750367

## GENERAL INFORMATION:

APPLICANT: Lawrence C. B. Chan  
TITLE OF INVENTION: HUMAN AND MOUSE VERY LOW  
DENSITY LIPOPROTEIN RECEPTORS  
TITLE OF INVENTION: AND METHODS FOR USE OF SUCH  
RECEPTORS  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LYON & LYON  
STREET: 611 West Sixth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90017

## COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: IBM MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/149,103

## FILING DATE:

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below:  
APPLICATION NUMBER: none

## FILING DATE:

## ATTORNEY/AGENT INFORMATION:

NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 204/052  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510

## INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 3330 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-149-103-1

Query Match 1.5%; Score 25; DB 1; Length 3330;  
Best Local Similarity 100.0%; Pred. No. 0.074;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1337 acggcgccgagcgccgagcg 1361  
|||||  
DB 67 ACGGCGCGCGCGCGCGCGCGG 91

RESULT 12

US-08-451-883-1  
Sequence 1, Application US/08451883  
Patent No. 5798209

## GENERAL INFORMATION:

APPLICANT: Lawrence C.B. Chan  
TITLE OF INVENTION: HUMAN AND MOUSE VERY LOW DENSITY  
LIPOPROTEIN RECEPTORS AND METHODS FOR  
TITLE OF INVENTION: USE OF SUCH RECEPTORS

## NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:  
ADDRESSEE: LYON & LYON  
STREET: 633 West Fifth Street, Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066

## COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: IBM MS-DOS (Version 6.22)  
SOFTWARE: WordPerfect (Version 5.1)

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/451,883

## FILING DATE:

## CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below:  
APPLICATION NUMBER: 08/149,103

FILING DATE: No. 5798209ember 8, 1993

## ATTORNEY/AGENT INFORMATION:

NAME: Knight, Matthew W.  
REGISTRATION NUMBER: 36,846  
REFERENCE/DOCKET NUMBER: 212/268  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510

## INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 3330 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-451-883-1

Query Match 1.5%; Score 25; DB 1; Length 3330;  
Best Local Similarity 100.0%; Pred. No. 0.074;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1337 acggcgccgagcgccgagcg 1361  
|||||  
DB 67 ACGGCGCGCGCGCGCGCGCGG 91

RESULT 13

US-08-393-734-1  
Sequence 1, Application US/08393734  
Patent No. 5652224

## GENERAL INFORMATION:

APPLICANT: Wilson, James M.  
APPLICANT: Kozarsky, Karen F.  
APPLICANT: Strauss, Jerome F.  
TITLE OF INVENTION: Methods and Compositions for Gene  
TITLE OF INVENTION: Therapy for the Treatment of Defects in Lipoprotein  
TITLE OF INVENTION: Metabolism  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Cntr., PO Box 457  
CITY: Spring House  
STATE: Pennsylvania  
COUNTRY: USA

ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

[illegible]

RESULT 15  
 US-08-393-734-3  
 : Sequence 3, Application US/08393734  
 : Patent No. 5652224  
 :  
 : GENERAL INFORMATION:  
 : APPLICANT: Wilson, James M.  
 : APPLICANT: Kozarsky, Karen F.  
 : APPLICANT: Straus, Jerome F.  
 : TITLE OF INVENTION: Methods and Compositions for Gene  
 : TITLE OF INVENTION: Therapy for the Treatment of Defects in Lipoprotein  
 : TITLE OF INVENTION: Metabolism  
 : NUMBER OF SEQUENCES: 8  
 :  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Howson and Howson  
 : STREET: Spring House Corporate Cntr., PO Box 457  
 : CITY: Spring House  
 : STATE: Pennsylvania  
 : COUNTRY: USA  
 : ZIP: 19477  
 :  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: Patent in Release #1.0, Version #1.30  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/393,734  
 : FILING DATE:  
 :  
 : CLASSIFICATION: 424  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Bak, Mary E.  
 : REGISTRATION NUMBER: 31,215  
 : REFERENCE/DOCKET NUMBER: UPMH1254USA  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 215-540-9200  
 : TELEFAX: 215-540-5818  
 : INFORMATION FOR SEQ ID NO: 3:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 9592 base pairs  
 : TYPE: nucleic acid  
 : STRANDEDNESS: double  
 : TOPOLOGY: unknown





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 6, 2001, 18:25:51 ; Search time 86.45 Seconds  
(without alignments)  
7222.110 Million cell updates/sec

Title: US-09-292-862-1

Perfect score: 1662  
Sequence: 1 atcgagcgcgactactccgt.....acgactgtagcaattttga 1662

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 480022 seqs, 187831343 residues

Word size: 0

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

N\_Geneseq\_36:\*

- 1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT:\*
- 2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT:\*
- 3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT:\*
- 4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT:\*
- 5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT:\*
- 6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT:\*
- 7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT:\*
- 8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT:\*
- 9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT:\*
- 10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT:\*
- 11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT:\*
- 12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT:\*
- 13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT:\*
- 14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT:\*
- 15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT:\*
- 16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT:\*
- 17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT:\*
- 18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT:\*
- 19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT:\*
- 20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT:\*
- 21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1662	100.0	1976	20	Human FREAC3 prote
2	1115	67.1	1662	20	Human forkhead tra
3	1115	67.1	1662	20	Human FKHL7 coding
4	1115	67.1	1662	20	Human forkhead tra
5	1115	67.1	1662	20	Human FKHL7 coding
6	68	4.1	2106	20	Human FKHL7 coding
7	47	2.8	1155	14	Brain factor-3 par
8	28	1.7	1860	14	Brain factor-2
9	27	1.6	640	18	Human calpain smal
10	27	1.6	804	18	Human calpain smal
11	27	1.6	2312	20	Chicken Sox1 cDNA
12	27	1.6	3658	20	Human 3-OST-4 enco

13	27	1.6	4045	20	X37251	Human 3-OST-4 5' p
14	26	1.5	410	21	A31290	Plant microsatelli
15	25	1.5	286	21	A31736	Plant microsatelli
16	25	1.5	316	21	A31188	Plant microsatelli
17	25	1.5	338	21	A31530	Plant microsatelli
18	25	1.5	377	21	A31325	Plant microsatelli
19	25	1.5	400	21	A31240	Plant microsatelli
20	25	1.5	434	21	A31517	Plant microsatelli
21	25	1.5	1308	17	T31602	Cartilage-derived
22	25	1.5	2375	21	A34740	Human adenosine re
23	25	1.5	3306	11	Q05620	Expressible sequen
24	25	1.5	3330	16	O88687	Human very low den
25	25	1.5	3505	20	X84667	Human KDR signal t
26	25	1.5	3656	17	T36751	VDL receptor gene
27	25	1.5	4475	21	A34741	Human adenosine re
28	25	1.5	5482	21	A09255	Human alpha-2-delt
29	25	1.5	8438	15	O73500	DNA encoding pseud
30	25	1.5	9592	17	T36752	Adenovirus vector
31	24	1.4	25	15	O55856	Fragile X probe.
32	24	1.4	25	16	O85271	Probe for Fragile
33	24	1.4	25	20	X05267	Fragile X chromoso
34	24	1.4	51	19	V68355	Adapter primer oli
35	24	1.4	67	16	O79642	Nucleotide sequenc
36	24	1.4	75	16	O79643	Longest observed a
37	24	1.4	102	10	N91106	Epitope to poliovi
38	24	1.4	102	10	N94572	Epitope to foot an
39	24	1.4	113	21	A35570	Microsatellite nuc
40	24	1.4	119	19	V16155	Microsatellite mar
41	24	1.4	126	21	A31060	Plant microsatelli
42	24	1.4	131	19	V16156	Microsatellite mar
43	24	1.4	149	10	N94576	Epitope to Malaria
44	24	1.4	152	19	V16158	Microsatellite mar
45	24	1.4	171	21	A31044	Plant microsatelli

## ALIGNMENTS

RESULT 1	
ID Z28095	standard; cDNA; 1976 BP.
AC Z28095;	
DT 31-JAN-2000	(first entry)
XX Human FREAC3	protein encoding cDNA.
DE FREAC3 gene; eye disease; eye developmental defect; therapy: mutation;	
KW glaucoma; anterior segment dysgenesis; ss.	
KX	
OS Homo sapiens.	
XX	
FH Key	Location/Qualifiers
FT CDS	159..1820
FT	/*tag= a
FT	/product= "FREAC3"
FT	403
FT	/*tag= b
FT	/note= "a G to C transversion mutation at this position can result in a Ser827Trp mutation in helix 1 of the FREAC3 forkhead domain"
FT	419
FT	/*tag= c
FT	/note= "a G to C transversion mutation at this position can result in a Ile877Met mutation in helix 1 of the FREAC3 forkhead domain"
PN WO954493-A2.	
XX	
PD 28-OCT-1999.	
XX	
FT 16-APR-1999;	99WO-IB01024.











[illegible]

RESULT	5
Z31671	
ID	Z31671 standard; DNA; 3946 BP.
XX	
AC	Z31671;
XX	
DT	17-JAN-2000 (first entry)
XX	
DE	Human FKHL7 coding sequence.
XX	
KW	FKHL7; human; forkhead transcription factor gene; diagnosis; therapy;
KW	congenital heart disease; ss.
XX	
OS	Homo sapiens.

PN WO952415-A2.  
 XX  
 PD 21-OCT-1999.  
 XX  
 PF 14-APR-1999; 99WO-US08159.  
 XX  
 PR 15-APR-1998; 98US-0081870.  
 PR 22-MAY-1998; 98US-0083351.  
 XX  
 PA (IOWA ) UNIV IOWA RES FOUND.  
 XX  
 XX Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;  
 DR WPI; 1999-620257/53.  
 DR P-PSDB; Y43260.  
 XX  
 PT New isolated human forkhead transcription factor gene, FKHL7, used to  
 PT develop products for the diagnosis, prognosis, monitoring, prevention  
 XX or treatment of congenital heart disease  
 PS  
 PS Claim 11; Fig 1; 98pp; English.  
 XX  
 CC This sequence encodes the human forkhead transcription factor gene,  
 CC designated FKHL7, of the invention. FKHL7 can be used in a novel meth  
 CC for treating or preventing the development of a congenital heart dise  
 CC (CHD) in a subject. The FKHL7 sequences can be used for diagnosis,  
 CC prognosis, monitoring, prevention and treatment of CHD. They can also  
 CC used for the production of transgenic animals and drug screening.  
 XX  
 XX Sequence 3946 BP; 905 A; 1155 C; 1053 G; 833 T; 0 other:

[illegible]



```

RESULT      7
050633 ID Q50633 standard; cDNA; 1155 BP.
XX AC Q50633;
XX DT 02-JUN-1994 (first entry)
XX DE Brain factor-3 partial sequence.
KW KW Brain factor; BF-1; BF-2; BF-3; DNA binding domain;
KM brain transcription factor; diagnosis; tumour; cancer; probe;
XX telencephalon; ss.
OS Rattus rattus.
XX PN W09323430-A.
XX PD 25-NOV-1993.
XX PF 30-APR-1993; 93WO-US04102.
XX PR 13-MAY-1992; 92US-0882292.
XX PA (SLOK ) SLOAN KETTERING INST CANCER.
XX PI Lai E, Tao W;
XX DR WPI; 1993-386481/48.
XX PT Isolated, animal nucleic acids encoding brain transcription
PT factors - useful for diagnosis and treatment of abnormal brain
PT factor synthesis in tumour tissue from animals and utilised as
PT probe
XX PS Claim 3; Page 63; 96pp; English.
XX CC The brain factors are transcription factors. Abnormal expression of
CC BF-1 in telencephalon-derived tissue or tumour tissue can be
CC diagnosed. BF Factor or protein can be used to correct defective
CC synthesis of BF.
XX SO Sequence 1155 BP; 211 A; 414 C; 367 G; 163 T; 0 other;

Query Match          2.8%; Score 47; DB 14; Length 1155;
Best Local Similarity 100.0%; Pred.No.5.6e-11;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 415 gtgcgcgcgaagcaagaacgccggccaaggcagctactgagacct 461
   |||||||
Db 682 gtgcgcgcgaagcaagaacgccggccaaggcagctactgagacct 728

RESULT      8
050632 ID Q50632 standard; cDNA; 1860 BP.
XX AC Q50632;
XX DT 02-JUN-1994 (first entry)
XX DE Brain factor-2.
XX KW Brain factor; BF-1; BF-2; BF-3; DNA binding domain;
KM brain transcription factor; diagnosis; tumour; cancer; probe;
XX telencephalon; ss.
OS Rattus rattus.
XX PN W09323430-A.
XX PD 25-NOV-1993.
XX PF 30-APR-1993; 93WO-US04102.
XX PR 13-MAY-1992; 92US-0882292.
XX PA (SLOK ) SLOAN KETTERING INST CANCER.
XX PI Lai E, Tao W;
XX DR WPI; 1993-386481/48.
XX PT Isolated, animal nucleic acids encoding brain transcription
PT factors - useful for diagnosis and treatment of abnormal brain
PT factor synthesis in tumour tissue from animals and utilised as
PT probe
XX PS Claim 3; Page 63; 96pp; English.
XX CC The brain factors are transcription factors. Abnormal expression of
CC BF-1 in telencephalon-derived tissue or tumour tissue can be
CC diagnosed. BF Factor or protein can be used to correct defective
CC synthesis of BF.
XX SO Sequence 1155 BP; 211 A; 414 C; 367 G; 163 T; 0 other;

```

[illegible]







```

ID A31290 standard; DNA: 410 BP.
XX
XX A31290;
AC
XX 05-JUL-2000 (first entry)
DT
XX
XX Plant microsatellite marker #251.
DE
XX
XX Plant microsatellite sequence; core repeat sequence; detection; probe;
KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
KW variety identification; genetic variability evaluation; primer; ss.
XX
XX Eucalyptus grandis.
OS
XX
XX WO9967421-A1.
PN
XX
XX 29-DEC-1999.
PD
XX
XX 25-JUN-1999; 99WO-NZ00092.
PF
XX
XX 25-JUN-1999; 98US-0105307.
PR
XX
XX (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX Havukkala IJ, Bloksberg LN, Glenn M;
PI
XX
XX WPI; 2000-116958/10.
DR
XX
XX New plant microsatellite markers and associated flanking species for
PT the detection of polymorphic genetic markers -
PS
XX
XX Claim 1; Page 150; 392pp; English.
SQ
XX
XX Sequences A31040-A32093 represent novel plant microsatellite sequences
CC and associated flanking species. The sequences comprise a central core
CC repeat sequence, especially selected from the sequences A32094-A32096
CC with left and right flanking sequences. The polynucleotide sequences
CC can be used in the detection of DNA polymorphisms, in genome mapping,
CC in physical mapping, in positional cloning of genes, in variety
CC identification and in evaluation of genetic variability within and
CC between plant tissues, populations, cultivars, species and species
CC groups. They may also be used to design hybridization probes for
CC oligonucleotide fingerprinting and library screening and to design
CC primers for microsatellite-primed PCR. Microsatellite markers are
CC useful to locate specific economically useful genes in plant genomes.
XX
XX
SQ Sequence 410 BP; 61 A; 164 C; 89 G; 96 T; 0 other;

Query Match 1.6%; Score 26; DB 21; Length 410;
Best Local Similarity 100.0%; Pred. No. 0.047;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1339 ggcggcgccggcgccggcgccggcg 1364
    ||||||||||||||||||||||||||||
DB 362 ggcggcgccggcgccggcgccggcg 337

RESULT 15
A31736
ID A31736 standard; DNA: 286 BP.
XX
XX A31736;
AC
XX
XX 05-JUL-2000 (first entry)
DT
XX
XX Plant microsatellite marker #697.
DE
XX
XX Plant microsatellite sequence; core repeat sequence; detection; probe;
KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
KW variety identification; genetic variability evaluation; primer; ss.
XX

```

```

OS Eucalyptus grandis.
XX
XX WO9967421-A1.
PN
XX
XX 29-DEC-1999.
PD
XX
XX 25-JUN-1999; 99WO-NZ00092.
PF
XX
XX 25-JUN-1999; 98US-0105307.
PR
XX
XX (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX Havukkala IJ, Bloksberg LN, Glenn M;
PI
XX
XX WPI; 2000-116958/10.
DR
XX
XX New plant microsatellite markers and associated flanking species for
PT the detection of polymorphic genetic markers -
PS
XX
XX Claim 1; Page 282; 392pp; English.
SQ
XX
XX Sequences A31040-A32093 represent novel plant microsatellite sequences
CC and associated flanking species. The sequences comprise a central core
CC repeat sequence, especially selected from the sequences A32094-A32096
CC with left and right flanking sequences. The polynucleotide sequences
CC can be used in the detection of DNA polymorphisms, in genome mapping,
CC in physical mapping, in positional cloning of genes, in variety
CC identification and in evaluation of genetic variability within and
CC between plant tissues, populations, cultivars, species and species
CC groups. They may also be used to design hybridization probes for
CC oligonucleotide fingerprinting and library screening and to design
CC primers for microsatellite-primed PCR. Microsatellite markers are
CC useful to locate specific economically useful genes in plant genomes.
XX
XX
SQ Sequence 286 BP; 72 A; 74 C; 91 G; 49 T; 0 other;

Query Match 1.5%; Score 25; DB 21; Length 286;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1338 cggcgcgccggcgccggcgccggcg 1362
    ||||||||||||||||||||||||||||
DB 98 cggcgcgccggcgccggcgccggcg 122

```

Search completed: February 6, 2001, 21:01:11  
Job time: 9320 sec

Wed Feb 7 14:05:53 2001

us-09-292-862-1.rng

---

Page 12

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 6, 2001, 18:01:15 ; Search time 1965.59 Seconds  
(without alignments)  
4327.302 Million cell updates/sec

Title: US-09-292-862-1  
Perfect score: 1662  
Sequence: 1 atgcagcgcgcctactccgt.....acgactctagcaagtttga 1662

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1118133 seqs, 2558875100 residues

Word size: 0

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_ph:\*  
6: gb\_pl1:\*  
7: gb\_pl2:\*  
8: gb\_pr1:\*  
9: gb\_pr2:\*  
10: gb\_pr3:\*  
11: em\_fun:\*  
12: em\_hum1:\*  
13: em\_hum2:\*  
14: em\_in:\*  
15: em\_om:\*  
16: em\_or:\*  
17: em\_ov:\*  
18: em\_pat:\*  
19: em\_ph:\*  
20: em\_pl:\*  
21: em\_ro:\*  
22: em\_sts:\*  
23: em\_sy:\*  
24: em\_un:\*  
25: em\_vl:\*  
26: gb\_htg1:\*  
27: gb\_htg2:\*  
28: gb\_in1:\*  
29: gb\_in2:\*  
30: em\_ba1:\*  
31: em\_ba2:\*  
32: em\_hum3:\*  
33: em\_hum4:\*  
34: gb\_pr4:\*  
35: gb\_htg3:\*  
36: gb\_htg4:\*  
37: gb\_htg5:\*  
38: gb\_htg6:\*  
39: gb\_htg7:\*  
40: em\_htg2:\*  
41: em\_htg3:\*  
42: em\_htg3:\*  
43: em\_hum5:\*

44: gb\_pl3:\*  
45: gb\_pr5:\*  
46: gb\_htg8:\*  
47: gb\_htg9:\*  
48: gb\_htg10:\*  
49: gb\_htg11:\*  
50: gb\_htg12:\*  
51: gb\_htg13:\*  
52: gb\_htg14:\*  
53: gb\_in3:\*  
54: gb\_htg15:\*  
55: gb\_htg16:\*  
56: gb\_htg17:\*  
57: em\_htg4:\*  
58: em\_htg5:\*  
59: em\_htg6:\*  
60: em\_htg7:\*  
61: em\_hum5:\*  
62: gb\_htg18:\*  
63: gb\_htg19:\*  
64: gb\_ba3:\*  
65: em\_htg8:\*  
66: em\_htg9:\*  
67: em\_htg10:\*  
68: gb\_pr6:\*  
69: gb\_pr7:\*  
70: gb\_htg20:\*  
71: gb\_htg21:\*  
72: gb\_htg22:\*  
73: gb\_htg23:\*  
74: gb\_ro:\*  
75: gb\_sts1:\*  
76: gb\_sts2:\*  
77: gb\_sy:\*  
78: gb\_un:\*  
79: gb\_vl1:\*  
80: gb\_vl2:\*  
81: gb\_pat1:\*  
82: gb\_pat2:\*  
83: em\_htg0:\*  
84: gb\_htg24:\*  
85: gb\_pr8:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1662	100.0	1662	10	AF078096	AF078096 Homo sapi
2	1319	79.4	104729	45	HS118B18	AL034344 Human DNA
3	1217	73.2	3946	10	AF048693	AF048693 Homo sapi
4	986	59.3	160855	35	AC011579	AC011579 Homo sapi
5	986	19.1	318	69	HSU13221	U13221 Human forkh
6	236	14.2	331	85	HOMFORKHOC	L12143 Human Droso
7	152	9.1	249	44	HS116B11F	Z56386 H.sapiens C
8	74	4.5	334	74	MMFKH1A	X71939 M.musculus
9	74	4.5	355	74	MUSPRKHDA	L10406 Mouse forkh
10	74	4.5	3579	74	AF045017	AJ223298 Mus muscu
11	74	4.5	3637	74	AF045017	AF045017 Mus muscu
12	68	4.1	3289	69	HSMPH1	Y08223 H.sapiens M
13	68	4.1	170431	27	AC009108	AC009108 Homo sapi
14	48	2.9	303	74	RATHFH3BF	L13193 Rattus norv
15	47	2.8	1155	81	AR083461	AR083461 Sequence
16	47	2.8	1942	74	S63607	S63607 MPR-1-mesen
17	47	2.8	2323	74	MMKRFH1	X92499 M.musculus
18	47	2.8	2712	74	MMKRFH1	X74040 M.musculus
19	47	2.8	6021	74	MMKRFH1	Y08222 M.musculus
20	45	2.7	318	69	HSU13225	U13225 Human forkh
21	42	2.5	180	4	GGU37277	U37277 Gallus gall





[illegible]

[illegible][illegible]

[illegible]

QY	1140	cggggggagccggggggggcgcgggcgggcgccggagctaccatctgcaacctgcaagccat	1199
Db	1614	cggggggggccggggggggcgggcgggcgccggcgccggagcttaccactgcacmcttgcagaccat	1673
QY	1200	gaagccctgtacagcgccagcgcgagcgcgagggccaacttgcaagggcgccccggggggcgcggg	1259
Db	1674	gaagccctgtacagcgggcgggcgagcgccggggggccactttccagggcgccccggggggcgcggg	1733
QY	1260	cgagctcgagcgctgtgagcaagaccctgtccgcactactctctgtccctggtlcaacagcaagca	1319
Db	1734	cggacttggccgggggcaacacccctgcggccgagctacttctgcctccgggtaccagacagag	1793
QY	1320	ctcgtctgtcccttgagtlcaacgcgcgagcgcgagcgcgagcgcgaggggaagccagagggcgcg	1379
Db	1794	ctcgtctgtctcccttgagttcaacgagcgggcgggcgggcgggcgggggggagggccagagagccgg	1853
QY	1380	ccaccacaccttcgggcccacccaagggcgcgctctcaactctgtgtactgaaaccaagcgcgagcg	1439
Db	1854	ccaccacaccttcctggcgcccaaccaaagggcgcccttcaacttctgtgtacttgcgaaccagggggcg	1513
QY	1440	agaccttggccaacttggcgagcgcgcgcgcgcgcgcgcgcgcgcgagcgagcgagcgagcgagcga	1499
Db	1914	agaccttggggccactttggccaagcgcgggcgggcgggcgggcgggcgggcgagggcttaccggggcca	1973
QY	1500	gcagagaagacttccaactcggltgggggaagatgttcgaatccaagagatcgagcttgaacaa	1559
Db	1974	gcagcagaacttccaactcgggttgggggaagatgttgcagatccaagagatcgagcttgaacaa	2033
QY	1560	ctctccagtggaacgggaatagtagctgttcaaaatggccttcccttccaagcagatctctgtga	1619
Db	2034	ctcttccagtggaacgggaatagtagctgttcaaaatggccttcccttccaagcagatctctgtga	2093
QY	1620	ccgcaagcttcggagccttctgtcttcaagacgtgtgcaagatttttga 1662	
Db	2094	ccgcaagcttcggagccttctgtcttcaagacgtgttgcattacgaccttgcacagatttttga 2136	

RESULT	4
LOCUS	AC011579/c
DEFINITION	
ACCESION	AC011579    160855 bp     DNA                HTG           20-SEP-2000
VERSION	AC011579
SOURCE	Homo sapiens clone RP11-13f18, WORKING DRAFT SEQUENCE, 53 unordered pieces.
ORGANISM	human. Homo sapiens Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 160855) Birren,B., Linton,L., Nushbaum,C. and Lander,E. Unpublished
AUTHORS	2 (bases 1 to 160855)
TITLE	Birren,B., Linton,L., Nushbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,I., Bouknighter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collumore,A., Cooke,P., Destrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferrella,T., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gallagher,J., Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczy,J., Lien,C., Locke,C., MacDonald,P., Marquis,N., McMan,P., McCutck,A., McKernan,K., McDonald,I.J., Meldrin,M., Morrow,J., Naylor,U., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye-S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
JOURNAL	Submitted (07-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	On Sep 20, 2000 this sequence version replaced g1:g123848.



All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIRB

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: 13\_F\_18

Center clone name: 13\_F\_18

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 136346 bases at least Q40

Consensus quality: 148591 bases at least Q30

Consensus quality: 153141 bases at least Q20

Insert size: 166000; agarose-fp

Insert size: 155655; sum-of-ctrls

Quality coverage: 3.1 in Q20 bases; sum-of-ctrls

Quality coverage: 3.3 in Q20 bases; sum-of-ctrls

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 53 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1      4649: contig of 4649 bp in length
*      4650 4749: gap of 100 bp
*      4750 5802: contig of 1053 bp in length
*      5803 5902: gap of 100 bp
*      5903 7063: contig of 1161 bp in length
*      7064 7163: gap of 100 bp
*      7164 8774: contig of 1611 bp in length
*      8775 8874: gap of 100 bp
*      8875 10112: contig of 1238 bp in length
*      10113 10212: gap of 100 bp
*      10213 11283: contig of 1071 bp in length
*      11284 11383: gap of 100 bp
*      11384 12607: contig of 1224 bp in length
*      12608 12707: gap of 100 bp
*      12708 14285: contig of 1578 bp in length
*      14286 14385: gap of 100 bp
*      14386 15414: contig of 1029 bp in length
*      15415 15514: gap of 100 bp
*      15515 17693: contig of 2179 bp in length
*      17694 17793: gap of 100 bp
*      17794 18837: contig of 1044 bp in length
*      18838 18937: gap of 100 bp
*      18938 20536: contig of 1599 bp in length
*      20537 20636: gap of 100 bp
*      20637 35792: contig of 15156 bp in length
*      35793 35892: gap of 100 bp
*      35893 37876: contig of 1984 bp in length
*      37877 37976: gap of 100 bp
*      37977 40070: contig of 2094 bp in length
*      40071 40170: gap of 100 bp
*      40171 41423: contig of 1253 bp in length
*      41424 41523: gap of 100 bp
*      41524 42814: contig of 1291 bp in length
*      42815 42914: gap of 100 bp
*      42915 44991: contig of 2077 bp in length
*      44992 45091: gap of 100 bp
*      45092 46704: contig of 1613 bp in length
*      46705 46804: gap of 100 bp
*      46805 49273: contig of 2469 bp in length
*      49274 49373: gap of 100 bp
*      49374 51528: contig of 2155 bp in length
*      51529 51628: gap of 100 bp

```

```

*      51629 52804: contig of 1176 bp in length
*      52805 52904: gap of 100 bp
*      52905 55108: contig of 2204 bp in length
*      55109 55208: gap of 100 bp
*      55209 57532: contig of 2324 bp in length
*      57533 57632: gap of 100 bp
*      57633 59640: contig of 2008 bp in length
*      59641 59740: gap of 100 bp
*      59741 61382: contig of 1642 bp in length
*      61383 61482: gap of 100 bp
*      61483 64104: contig of 2622 bp in length
*      64105 64204: gap of 100 bp
*      64205 66859: contig of 2655 bp in length
*      66860 66959: gap of 100 bp
*      66960 68324: contig of 1365 bp in length
*      68325 68424: gap of 100 bp
*      68425 71212: contig of 2788 bp in length
*      71213 71312: gap of 100 bp
*      71313 74324: contig of 3012 bp in length
*      74325 74424: gap of 100 bp
*      74425 76988: contig of 2564 bp in length
*      76989 77088: gap of 100 bp
*      77089 78929: contig of 1841 bp in length
*      78930 79029: gap of 100 bp
*      79030 81494: contig of 2465 bp in length
*      81495 81594: gap of 100 bp
*      81595 82738: contig of 1144 bp in length
*      82739 82838: gap of 100 bp
*      82839 86549: contig of 3711 bp in length
*      86550 86649: gap of 100 bp
*      86650 90087: contig of 3438 bp in length
*      90088 90187: gap of 100 bp
*      90188 93458: contig of 3271 bp in length
*      93459 93558: gap of 100 bp
*      93559 97847: contig of 4289 bp in length
*      97848 97947: gap of 100 bp
*      97948 101934: contig of 3987 bp in length
*      101935 102034: gap of 100 bp
*      102035 106306: contig of 4272 bp in length
*      106307 106406: gap of 100 bp
*      106407 110682: contig of 4276 bp in length
*      110683 110782: gap of 100 bp
*      110783 115744: contig of 4962 bp in length
*      115745 115844: gap of 100 bp
*      115845 117155: contig of 1311 bp in length
*      117156 117255: gap of 100 bp
*      117256 123238: contig of 5983 bp in length
*      123239 123338: gap of 100 bp
*      123339 128170: contig of 4832 bp in length
*      128171 128270: gap of 100 bp
*      128271 133489: contig of 5219 bp in length
*      133490 133589: gap of 100 bp
*      133590 138250: contig of 4661 bp in length
*      138251 138350: gap of 100 bp
*      138351 144214: contig of 5864 bp in length
*      144215 144314: gap of 100 bp
*      144315 151760: contig of 7446 bp in length
*      151761 151860: gap of 100 bp
*      151861 158782: contig of 6922 bp in length
*      158783 158882: gap of 100 bp
*      158883 160198: contig of 1316 bp in length
*      160199 160298: gap of 100 bp
*      160299 160855: contig of 557 bp in length.

```

#### FEATURES

source

1. 160855

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="RP11-13F18"

/clone\_lib="RP11 Human Male BAC"

1. 4649

/note="assembly\_fragment"

clone\_end:Sp6

vector\_side:left"

misc\_feature

Query Match	59.38;	Score 986;	DB 35;	Length 160855;
Best Local Similarity	99.98;	Pred. No. 0;		
Matches 1036; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0

[illegible]

DEFINITION	Human forkhead protein FREAC-3 mRNA, partial cds.	PRI	13-MAR-1996
ACCESSION	U13221		
VERSION	013221.1	GI:563161	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	1 (bases 1 to 318) Pierrou,S., Helligvist,M., Samuelsson,L., Enerback,S. and Carlsson,P.		
TITLE	Cloning and characterization of seven human forkhead proteins: binding site specificity and DNA bending		
JOURNAL	EMBO J. 13 (20), 5002-5012 (1994)		
MEDLINE	95045392		
REFERENCE	2 (bases 1 to 318) Carlsson,P.		
AUTHORS	Direct Submission		
TITLE	Submitted (10-AUG-1994) Peter Carlsson, Molecular Biology, Goteborg University, Medicinaregatan 9C, Goteborg S-413 90, Sweden		
JOURNAL	Location/Qualifiers		
FEATURES	1..318		
source	/organism="Homo sapiens" /db_xref="taxon:9606"		
CDS	<1..>318 /codon_start=1 /product="FREAC-3" /protein_id="AA92038.1" /db_xref="GI:563162" /translation="PKDMVKRPYSIALITMAIONAPDKKITINGIYOFIMDFEPPYRDNKGQWNSIRHNLSLNECFVKVPRDDKKRGKGSYWTLPDPSYMFENGSFLRRRRFKKKD"		
misc_feature	16..315 /note="forkhead-motif"		
BASE COUNT	80 a 107 c 85 g 46 t		
ORIGIN			
Query Match	19.1%; Score 318; DB 69; Length 318;		
Best Local Similarity	100.0%; Pied No.1.3e-125;		
Matches 318; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
0y	217 cccaagacatggtggaagccgcctatagctacatgcgcgtcatcaccatgagcatccag		276

Db 1 CCCAAGCAGATGTTGAAGCCGCTATAGTACATCGCGCTCATCAGCATGGCATCCAG 60

Qy 277 aagcccccagacaagaatcatccctgaacgagcatctaccagttcatctgacgcttc 336  
|||||  
Db 61 AACGCCCGGACAGAAATCATCCCTGAGCGCATCTACAGATTGATCATGAGCCGCTTC 120  
|||||

Qy 337 cccttctacggagacaagaacgaggtctgagacagcatccgcaaacctctctgc 396  
|||||  
Db 121 CCCTTACACGGGACACAGAGGCTGGCAGAACAGCATCCGCCACACCTCTGCTC 180  
|||||

Qy 397 aacgagttctgctcaagttgctgcgagcagacaagaagccgaggaaggaactatg 456  
|||||  
Db 181 AAGGATGCTTGTCTCAAGTGGCGGCGACGACAGAAAGCCGGGCAAGGCGAGTACTG 240  
|||||

Qy 457 agcgtgacccggaactctcaacacatgtctgagaagcagcttctctgagcgagcg 516  
|||||  
Db 241 ACGGTGACCGGACCTCTCAACATGTTCCAGAACGCGACCTCTGCGCGCGGCGG 300  
|||||

Qy 517 cgtctcaagaagaagac 534  
|||||  
Db 301 CGCTTCAAGAAAGAGAC 318  
|||||

RESULT 6  
HUMFORRHOC 331 bp mRNA PRI 04-AUG-1993  
LOCUS Human Drosophila forkhead homologue mRNA.  
DEFINITION L12143  
ACCESSION L12143.1 GI:347525  
VERSION  
KEYWORDS forkhead homolog.  
SOURCE Homo sapiens (library: lambda gt11) blood cDNA to mRNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 331)  
Hromas, R.A., Moore, J.L., Johnston, T.C., Socha, C. and Klemenz, M.J.  
Drosophila forkhead homologues are expressed in a  
lineage-restricted manner in human hematopoietic cells  
Blood 81, 2854-2859 (1993)  
JOURNAL 93271467  
MEDLINE  
FEATURES  
source location/Qualifiers  
1. .331  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_type="chronic myelogenous leukemia"  
/tissue\_type="blood"  
/tissue\_1lb="lambda gt11"  
BASE COUNT 85 a 113 c 87 g 46 t  
ORIGIN

Query Match 14.2% Score 236; DB 85; Length 331;  
Best Local Similarity 99.7%; Pred. No. 1.7e-90;  
Matches 286; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 232 aagcgcctatagctacatcggtctatccatcaggtccatccgagcccgagacaag 291  
|||||  
Db 31 AAGCGGCTATAGTACATCGCGCTCATCAGCATGGCATCCAGAGCCCGGACAG 90  
|||||

Qy 292 aagtcacccctgaaggaacatctacagttcatcagagcgttcccttctacagagc 351  
|||||  
Db 91 AAGATCACCCAGAACGGGACATCTACAGATTCATGAGCGCTTCCCTTACCGGAGC 150  
|||||

Qy 352 aacaagcagaggtctgagacaacagcatccgcaaacctctgctcaagagaggtctg 411  
|||||  
Db 151 AACAGCAGGCGCTGGCAGAACAGCATCCGCCACACCTCTCTCAAGAGGCTTCTGTC 210  
|||||

Qy 412 aagtgctgcgagcagacaagaacccgaggaagagcagctactgagcgtctgagccgagc 471  
|||||  
Db 211 AAGGTGCGCGCAGAGAACAGACCCGGGCAAGGCGAGCTACTGAGACGCTGGACCGGAGC 270  
|||||

Qy 472 tctctacaatgctcgaagaagcagcttctctgagcgagcgagcg 518  
|||||

Db 271 TCCTACACATGTTGAGAACGCGAGCTTCTCGCGGCGGCGCGGCG 317

RESULT 7  
HS116B11F/c 249 bp DNA PRI 18-OCT-1995  
LOCUS H.sapiens Cpg island DNA genomic MseI fragment, clone 116b11,  
DEFINITION forward read cpg116b11.fita.  
ACCESSION 256386  
VERSION 256386.1 GI:1027617  
KEYWORDS Cpg island; genomic MseI fragment.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 249)  
Macdonald, M., Huckle, E., Wilkinson, P. and Micklem, G.  
Direct Submission  
Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1RQ, England. E-mail contact: humquery@sanger.ac.uk  
JOURNAL 94282070  
REFERENCE 2 (bases 1 to 249)  
Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.  
Purification of Cpg islands using a methylated DNA binding column  
Nat. Genet. 6 (3), 236-244 (1994)  
MEDLINE 94282070  
COMMENT Vector: pGEM-5Zf(-)  
Clones are available from the UK MRC Human Genome Mapping Project  
Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:  
http://www.hgmp.mrc.ac.uk/ for details  
or contact: biolnephgmp.mrc.ac.uk.

FEATURES  
source location/Qualifiers  
1. .249  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/sex="male"  
/dev\_stage="adult"  
/tissue\_type="blood"  
/clone\_1lb="CGI-1"  
/clone="116b11"  
BASE COUNT 48 a 52 c 66 g 81 t 2 others  
ORIGIN

Query Match 9.1%; Score 152; DB 45; Length 249;  
Best Local Similarity 100.0%; Pred. No. 1.9e-54;  
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1511 tcaactggtggtggagatcttcagttcacagagatcaggtctggaacaactctccagtga 1570  
|||||  
Db 249 TCCACTCGGTGCGGAGATGTTGAGATCAGAGAGATCGGCTTGAACAACCTCCAGTGA 190  
|||||

Qy 1571 acggaatagtagctgtcaaatgagcttcccttcagcagcagtcctgtaccgagctcg 1630  
|||||  
Db 169 ACGGGAATAGTAGCTGTCAATATGCGCTTCCCTTCCAGCGAGTCTGTACCGACGTCCG 130  
|||||

Qy 1631 gagcttctgcttaagactgttagcaagtttga 1662  
|||||  
Db 129 GAGCTTGTCTACGACTGTAGCAAGTTTGA 98  
|||||

RESULT 8  
MERKHA 334 bp DNA ROD 25-AUG-1995  
LOCUS M.musculus fkh-1 gene.  
DEFINITION X71939  
ACCESSION X71939.1 GI:311736  
VERSION DNA binding domain; fork head domain.  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 334)  
Kastner, K.H.

TITLE Direct Submission  
 JOURNAL Submitted (12-MAY-1993) K.H. Kaestner, German Cancer Center, Im  
 Neuenheimer Feld 280, Heidelberg 6900, FRG  
 REFERENCE 2 (bases 1 to 334)  
 AUTHORS Kaestner, K.H., Lee, K.H., Schlondorff, J., Hiemisch, H., Monaghan, A.P.  
 and Schütz, G.  
 TITLE Six members of the mouse forkhead gene family are developmentally  
 regulated  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90 (16), 7628-7631 (1993)  
 MEDLINE 93361500  
 FEATURES  
 source Location/Qualifiers  
 1..334  
 /organism="Mus musculus"  
 /strain="129"  
 /db\_xref="taxon:10090"  
 /dev\_stage="embryonic"  
 /cell\_type="stem cell"  
 1..334  
 /gene="fkh-1"  
 /note="DNA binding domain; forkhead domain"

gene 1..334  
 /note="DNA binding domain; forkhead domain"  
 CDS <1..>334  
 /gene="fkh-1"  
 /codon\_start=1  
 /protein\_id="CAA50741.1"  
 /db\_xref="GI:311737"  
 /db\_xref="MGI:95544"  
 /db\_xref="SWISS-PROT:O61572"  
 /translation="POPOKDMVKPVSIALITMAIONAPDKITLNGIYOFIMDRF  
 FYRDNKOGWMSIRHNSLNECFVKYPRDDKPKGKSYWTLDPDSYNMFENGSFLRR  
 RRRFKKKA"

BASE COUNT 85 a 109 c 88 g 52 t  
 ORIGIN

Query Match 4.5%; Score 74; DB 74; Length 334;  
 Best Local Similarity 100.0%; Pred. No. 4 6e-21;  
 Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 463 gaccggactctacacatgttcgagaagcagctctctgcgagcgcgcgttc 522  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 259 GACCCGGACTCTACACATGTTCGAGAAGCAGCTTCTCGCGCGCGCGCTTC 318  
 OY 523 aagaagaagcgc 536  
 ||||||||||||||||  
 Db 319 AAGAAGAAGCAGC 332

RESULT 9  
 MUSFRKHA 355 bp mRNA ROD 01-NOV-1993  
 LOCUS Mouse fork head related protein mRNA, partial cds.  
 DEFINITION L10406  
 ACCESSION L10406.1 GI:404758  
 VERSION developmentally regulated gene; fork head domain; fork head related  
 protein.  
 KEYWORDS Mus musculus 8.5 days post coltum CDNA to mRNA.  
 SOURCE Mus musculus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 355)  
 AUTHORS Sasaki, H. and Hogan, B.L.  
 TITLE Differential expression of multiple fork head related genes during  
 gastrulation and axial pattern formation in the mouse embryo  
 JOURNAL Development 118, 47-59 (1993)  
 MEDLINE 93387221  
 FEATURES  
 source Location/Qualifiers  
 1..355  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /dev\_stage="8.5 days post coltum"  
 <1..>355

CDS <1..>355

/note="fork head domain; putative"  
 /codon\_start=3  
 /product="fork head related protein"  
 /protein\_id="AA03159.1"  
 /db\_xref="GI:410056"  
 /translation="POPKDMVKPVSIALITMAIONAPDKITLNGIYOFIMDRF  
 FYRDNKOGWMSIRHNSLNECFVKYPRDDKPKGKSYWTLDPDSYNMFENGSFLRRR  
 RRRFKKKAETRTIG"  
 BASE COUNT 100 a 107 c 89 g 59 t  
 ORIGIN

Query Match 4.5%; Score 74; DB 74; Length 355;  
 Best Local Similarity 100.0%; Pred. No. 4.5e-21;  
 Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 463 gaccggactctacacatgttcgagaagcagctctctgcgagcgcgcgttc 522  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 259 GACCCGGACTCTACACATGTTCGAGAAGCAGCTTCTCGCGCGCGCGCTTC 314  
 OY 523 aagaagaagcgc 536  
 ||||||||||||||||  
 Db 315 AAGAAGAAGCAGC 328

RESULT 10  
 MUAJ3298 3579 bp DNA ROD 06-NOV-1998  
 LOCUS Mus musculus Fkh1/Mf1 gene, complete cds.  
 DEFINITION AJ223298  
 ACCESSION AJ223298.1 GI:3805940  
 VERSION Fkh1/Mf1 gene; transcription factor.  
 KEYWORDS house mouse.  
 SOURCE Mus musculus  
 ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 3579)  
 AUTHORS Hiemisch, H.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-JAN-1998) Hiemisch, H., Neurobiology, BASF-LYNX  
 Bioscience AG, Im Neuenheimer Feld 519, D-69120 Heidelberg, GERMANY  
 REFERENCE 2 (bases 1 to 3579)  
 AUTHORS Hiemisch, H., Schütz, G. and Kaestner, K.H.  
 TITLE The mouse Fkh1/Mf1 gene: cDNA sequence, chromosomal localization  
 and expression in adult tissues  
 JOURNAL Gene 220 (1-2), 77-82 (1998)  
 MEDLINE 98440371  
 FEATURES  
 source Location/Qualifiers  
 1..3579  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 1..816  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="13"  
 /dev\_stage="embryonic day 10 p.c."  
 /map="17.02cM"  
 /tissue\_type="embryo"  
 1..2161  
 /gene="Fkh1/Mf1"  
 <1..159  
 /gene="Fkh1/Mf1"  
 500..2161  
 /gene="Fkh1/Mf1"  
 /function="putative transcription factor"  
 /codon\_start=1  
 /product="Fkh1/Mf1 protein"  
 /protein\_id="CAI1239.1"  
 /db\_xref="GI:3805941"  
 /translation="MCARYSVSSPSNGVYVYLGSGSYRAAAAAGGTTAMPAPM  
 GYVSHPAHEQYPCSMARAYGPTPOPOKDMVKPVSIALITMAIONAPDKITLNGIYOFIMDRF  
 FYRDNKOGWMSIRHNSLNECFVKYPRDDKPKGKSYWTLDPDSYNMFENGSFLRRR  
 RRRFKKKAETRTIG"

gene promoter  
 CDS

MFENGSFLRRR

## source

OPPPVRIODITKTEGTCSPSPPLSPALGSGSATYKIESPDSSSSSSLSGSSPP  
 GSLSAPRLSLDAEPAPPPAPPPHSGSVNIMTSLGSPGSAALGSGLLA  
 SAASRRAGIAPPLALGAVSPGSSLYSPGSSSSAGSGGGGGGGSSSAG  
 TCGAATYHCNLOAMSLYAAERGCHLOGPAGGASAAVDPLPDYSLPPATSSSSSL  
 SHGGGGEASHPHSHOGRLTSWLNAGDGLHLASAAAAAAGYGGQOONPHSVR  
 EMEFESORIGLNNSPVNGSSCOMAFPASOSLRTSGAFYDCSKF"

BASE COUNT 802 a 1020 c 964 g 793 t  
 ORIGIN

Query Match 4.5%; Score 74; DB 74; Length 3579;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-21;  
 Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 463 gaccgcgactctacacatgltcgaagacgagcttcctcgcgcgcgcgccttc 522  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 962 GACCCGACCTCTACACATGTCGAGAACGCGACCTCTCGCGCGCGCGCTTC 1021  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 OY 523 aagaagaagagcgc 536  
 ||||||||||||||||  
 DB 1022 AAGAAGAAGAGCAGC 1035

## RESULT 11

AF045017 3637 bp mRNA ROD 25-JUN-1998  
 LOCUS AF045017 Mus musculus mesoderm/mesenchyme forkhead 1 (Mf1) mRNA, complete cds.  
 DEFINITION AF045017  
 ACCESSION AF045017 GI:3228521  
 VERSION AF045017.1 GI:3228521  
 KEYWORDS house mouse.  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 3637)  
 Kume,T., Deng,K.Y., Winfrey,V., Gould,D.B., Walter,M.A. and Hogan,B.L.M.  
 The forkhead/winged helix gene Mf1 is disrupted in the pleiotropic mouse mutation congenital hydrocephalus

JOURNAL Cell 93 (6), 985-996 (1998)  
 MEDLINE 98297351  
 REFERENCE 2 (bases 1 to 3637)  
 Kume,T.  
 Direct Submission  
 Submitted (27-JAN-1998) Cell Biology, Vanderbilt University Medical School, 1161 21st Ave., S., C-2310 MCN, Nashville, TN 37232-2175.  
 JOURNAL School, 1161 21st Ave., S., C-2310 MCN, Nashville, TN 37232-2175, USA

## FEATURES

source Location/Qualifiers  
 1..3637  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="13"  
 /map="between D13Mit3 and D13B1r9"  
 1..3637  
 /gene="Mf1"  
 550..2211  
 /gene="Mf1"  
 /note="transcription factor"  
 /product="mesoderm/mesenchyme forkhead 1"  
 /protein\_id="AAC24209.1"  
 /db\_xref="GI:3228522"  
 /translation="MQARYSVSSPNSIGVVPYLGEGSYRAAAAAAGGCTAMPAPM  
 /SVSHPAHAEQYFGSMARAVGPTTPOPKDMVKRPYSYIALTIMAIONAPDKKITLN  
 GIYOFIMDRPPYRDNKGQMSINRHLSLNECFYKVRDRDKKPKSGSYWTLDPDSYN

BASE COUNT 823 a 1030 c 985 g 799 t  
 ORIGIN

Query Match 4.5%; Score 74; DB 74; Length 3637;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-21;  
 Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 463 gaccgcgactctacacatgltcgaagacgagcttcctcgcgcgcgcgccttc 522  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 1012 GACCCGACCTCTACACATGTCGAGAACGCGACCTCTCGCGCGCGCGCTTC 1071  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 OY 523 aagaagaagagcgc 536  
 ||||||||||||||||  
 DB 1072 AAGAAGAAGAGCAGC 1085

## RESULT 12

HSMFH1 3289 bp DNA PRI 14-MAY-1997  
 LOCUS HSMFH1 H.sapiens MFH-1 gene.  
 DEFINITION Y08223  
 ACCESSION Y08223.1 GI:1869804  
 VERSION Y08223.1 GI:1869804  
 KEYWORDS mesenchyme fork head-1 protein; MFH-1 gene.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 3289)  
 Miura,N., Iida,K., Kakimura,H., Yang,X.L. and Sugiyama,T.  
 Isolation of the mouse (MFH-1) and human (FKHL 14) mesenchyme fork head-1 genes reveals conservation of their gene and protein structures

JOURNAL Genomics 41 (3), 489-492 (1997)  
 MEDLINE 97312712  
 REFERENCE 2 (bases 1 to 3289)  
 Miura,N.  
 Direct Submission  
 Submitted (18-SEP-1996) N. Miura, Akita University School of Medicine, Department of Biochemistry, 1-1-1 Hondo, Akita 010, JAPAN

## FEATURES

source Location/Qualifiers  
 1..3289  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 1197..2702  
 /gene="MFH-1"  
 1197..2702  
 /gene="MFH-1"  
 /product="Mesenchyme Fork Head-1"  
 /protein\_id="CA69400.1"  
 /db\_xref="GI:1869805"  
 /db\_xref="SWISS-PROT:Q99598"  
 /translation="MQARYSVSDPNALGVVPLISRONYYRAAGSYGMAAPMGVYSGH  
 /PEQYASAGMSRYAPYHHHQPAPKDLVKEPVSYIALTIMAIONAPDKKITLNGIYOFI  
 MDREPFYRKNKQGMONSIRHNLSCNCFYKVRDRDKFKGSGYSTLTPDDSNMGENS  
 FLRRRRFRKKDYKSEKEERAKLKEPPPAASGAPATPHLDARKEKRVKVTSEDA  
 SPALPVYTKVETLSPESALOGSPRSAASTPAGSPGSLPEHHAAAPNGLPEFSEVNH  
 TLRTSPGGELSGACGAGLVVPPRLAAYPAAPAAAGCPGACOGLEAGAGAGYCSMK  
 AMSLYTGAERPAHMCYPRALDEALSDHPGPTSPALSLAAGDEGALAATGHHNHQ  
 GHHPQAPPPAPAPQPTPQVORPDAANAAQAAWYLNHSGDLNHLPGHTFAAOQTFPVV  
 REMFNHRIQENSTLGEVSQNSCQLPYHSTPPLYNHAAPYSDCTKY"  
 BASE COUNT 639 a 1125 c 925 g 600 t  
 ORIGIN

Query Match 4.1%; Score 68; DB 69; Length 3289;  
Best Local Similarity 100.0%; Pred. No. 9.6e-19;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 aacaagcaggtgcaagaacagcattccgcaaacctctgcctcaagatgcttcgc 411  
|||||  
Db 1530 AACAGCAGGCGTGGCAGAACGATCCGCCAACACTCTGCTCAAGCATGCTTCGTC 1589

QY 412 aaggtgccc 419  
|||||  
Db 1590 AAGGTGCC 1597

## RESULT 13

LOCUS AC009108 170431 bp DNA HTG 02-SEP-2000  
DEFINITION Homo sapiens chromosome 16 clone RP11-46309, WORKING DRAFT  
SEQUENCE, 4 ordered pieces.  
AC009108  
AC009108.8 GI:9964740  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
TITLE 1 (bases 1 to 170431)  
JOURNAL Sequencing of Human Chromosome 16  
REFERENCE Unpublished  
AUTHORS 2 (bases 1 to 170431)  
TITLE DOE Joint Genome Institute.  
JOURNAL Direct Submission  
COMMENT Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Sep 2, 2000 this sequence version replaced gi:8575957.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov

Project Information  
Center Project Name: 584502  
Center clone name: RP11-46309  
-----

## Summary Statistics

Consensus quality: 167345 bases at least Q40  
Consensus quality: 169423 bases at least Q30  
Consensus quality: 169900 bases at least Q20  
Estimated insert size: 181440; agarose-ff estimation  
Estimated insert size: 170331; agarose-ff estimation  
Quality coverage: 10.63 in Q20 bases; agarose-ff estimation  
Quality coverage: 11.33 in Q20 bases; sum-of-contigs estimation  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
1 24820: contig of 24820 bp in length  
24821 24920: gap of unknown length  
24921 37862: contig of 12942 bp in length  
37863 37962: gap of unknown length  
37963 114671: contig of 76709 bp in length  
114672 114771: gap of unknown length  
114772 170431: contig of 55660 bp in length.  
Location/Qualifiers  
1. 170431  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"

BASE COUNT 41117 a 43130 c 42591 g 43290 t 303 others  
ORIGIN

Query Match 4.1%; Score 68; DB 27; Length 170431;  
Best Local Similarity 100.0%; Pred. No. 3.5e-19;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 aacaagcaggtgcaagaacagcattccgcaaacctctgcctcaagatgcttcgc 411  
|||||  
Db 94939 AACAGCAGGCGTGGCAGAACGATCCGCCAACACTCTGCTCAAGCATGCTTCGTC 94998

QY 412 aaggtgccc 419  
|||||  
Db 94999 AAGGTGCC 95006

## RESULT 14

LOCUS RATHFH3BF 303 bp mRNA ROD 26-JUL-1993  
DEFINITION Rattus norvegicus brain factor-3 (HFH-BF-3) mRNA, complete cds.  
AC009108  
AC009108.1 GI:310156  
KEYWORDS forkhead homolog; hepatocyte nuclear factor; transcription factor.  
SOURCE Rattus norvegicus (strain Sprague-Dawley) adult cDNA to mRNA.  
ORGANISM

REFERENCE 1 (bases 1 to 303)  
AUTHORS Lal,E., Prezioso,V.R., Tao,W., Chen,W.S. and Darnell,J.E.Jr.  
TITLE Hepatocyte nuclear factor 3a belongs to a gene family in mammals  
that is homologous to the Drosophila homeotic gene fork head  
(correction)  
JOURNAL Unpublished (1992)  
REFERENCE 2 (bases 1 to 303)  
AUTHORS Clevidence,D.E., Overdier,D.G., Tao,W., Qian,X., Pani,L., Lal,E.  
and Costa,R.H.  
TITLE Identification of nine tissue-specific transcription factors of the  
hepatocyte nuclear factor 3/forkhead DNA-binding domain family  
Proc. Natl. Acad. Sci. U.S.A. 90, 3948-3952 (1993)  
93248207  
Location/Qualifiers  
1. 303  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/dev\_stage="adult"  
<1..>303  
/standard\_name="HFH-BF-3"  
/codon\_start=1  
/function="DNA binding domain"  
/product="brain factor-3"  
/protein\_id="AAA41320.1"  
/db\_xref="GI:310157"  
/translation="LVKPPSYVALSTRATONAEKKITINGLYVQFIMDRFPYRENN  
QOMNSIRHNLSINCEFYKVRDRKKPEKSGSYWTLDPDSYNMFEENGSLRRRRFRKK"

## CDs

BASE COUNT 76 a 94 c 84 g 49 t  
ORIGIN

Query Match 2.9%; Score 48; DB 74; Length 303;  
Best Local Similarity 100.0%; Pred. No. 6.6e-10;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 481 atgttcgaagcagcagcattccgctgcgagcgcgagcgttcaagaag 528  
|||||  
Db 256 ATGTTGAGAGAGCGAGCTTCTGCGCGCGCGCGCTTCAGGAG 303

## RESULT 15



